NASA CR-61258

MARKOV CHAIN TECHNIQUES FOR PREDICTING THE MAXIMUM WIND IN THE MAXIMUM DYNAMIC PRESSURE REGION FOR LAUNCHING SPACE VEHICLES

For

NASA-GEORGE C. MARSHALL SPACE FLIGHT CENTER
Marshall Space Flight Center, Alabama

MARKOV CHAIN TECHNIQUES FOR PREDICTING THE MAXIMUM WIND IN THE MAXIMUM DYNAMIC PRESSURE REGION FOR LAUNCHING SPACE VEHICLES

By

Harold L. Crutcher and Nathaniel B. Guttman

Prepared under Government Order 76789 by

ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
Environmental Data Service
National Weather Records Center
Asheville, North Carolina

NASA - GEORGE C. MARSHALL SPACE FLIGHT CENTER

TABLE OF CONTENTS

ABSTRACT

The wind fields of the atmospheric circulation at times present formidable hazards to the launching of a space vehicle. Acceptable predictions of the maximum wind in the maximum dynamic pressure region for space vehicles over Cape Kennedy, Florida (10-15 km) are sought for 12-hour increments through the use of transition matrices of operating Markov chains. The concepts of information theory, entropy, and Markovity are presented.

Empirical transition matrices are examined for stationarity and order of Markovity. The Markov models are compared to those of persistence and climatology. For winter and summer seasons predictions are made from each model and then verified. The test data are randomly selected from a period different from that used to construct the models.

Problems encountered during the study and recommendations for future investigation are discussed.

I. INTRODUCTION

The wind fields of the atmospheric circulation at times present formidable hazards to the launching of a space vehicle. Though in a macroscopic sense the atmospheric flow through which a vehicle passes may be relatively smooth, the shear from one level to another may be such as to adversely affect the vehicle's operation. Turbulence also may be detrimental to the passage of the vehicle.

The wind fields nearly always are in an intensifying or dissipating stage. These are seldom in a steady state stage. The definition of steady state may be given in terms of the atmosphere itself but would be better cast in terms of the vehicle's interaction with the atmosphere. This, sometimes, is difficult to determine. Therefore, this study is restricted to the prediction of only one feature of the wind field, namely, the maximum wind in the space vehicular dynamic pressure region, which is considered here to be 10 through 15 km. This paper has been presented in part as an invited paper at the Conference on High Altitude Meteorology and Space Weather at Houston, Texas, on March 29-31, 1967.

The National Aeronautics and Space Administration, Marshall Space Flight Center, R-AERO-YT, Huntsville, Alabama, (NASA-MSFC-R-AERO-YT) in cooperation with the Environmental Science Services Administration, Environmental Data Service, National Weather Records Center, Asheville, North Carolina (ESSA-EDS-NWRC) is developing prediction procedures for features of the wind distributions at Cape Kennedy, Florida.

Previous unpublished reports treat:

    A.   The static prediction of the maximum wind from the surface through 27 km, [16]

B. The prediction of the wind profile from the surface through 27 km by means of multiple regression techniques [14], and

C. The prediction of the wind profile from the surface through 27 km by use of Markov transition processes [15].
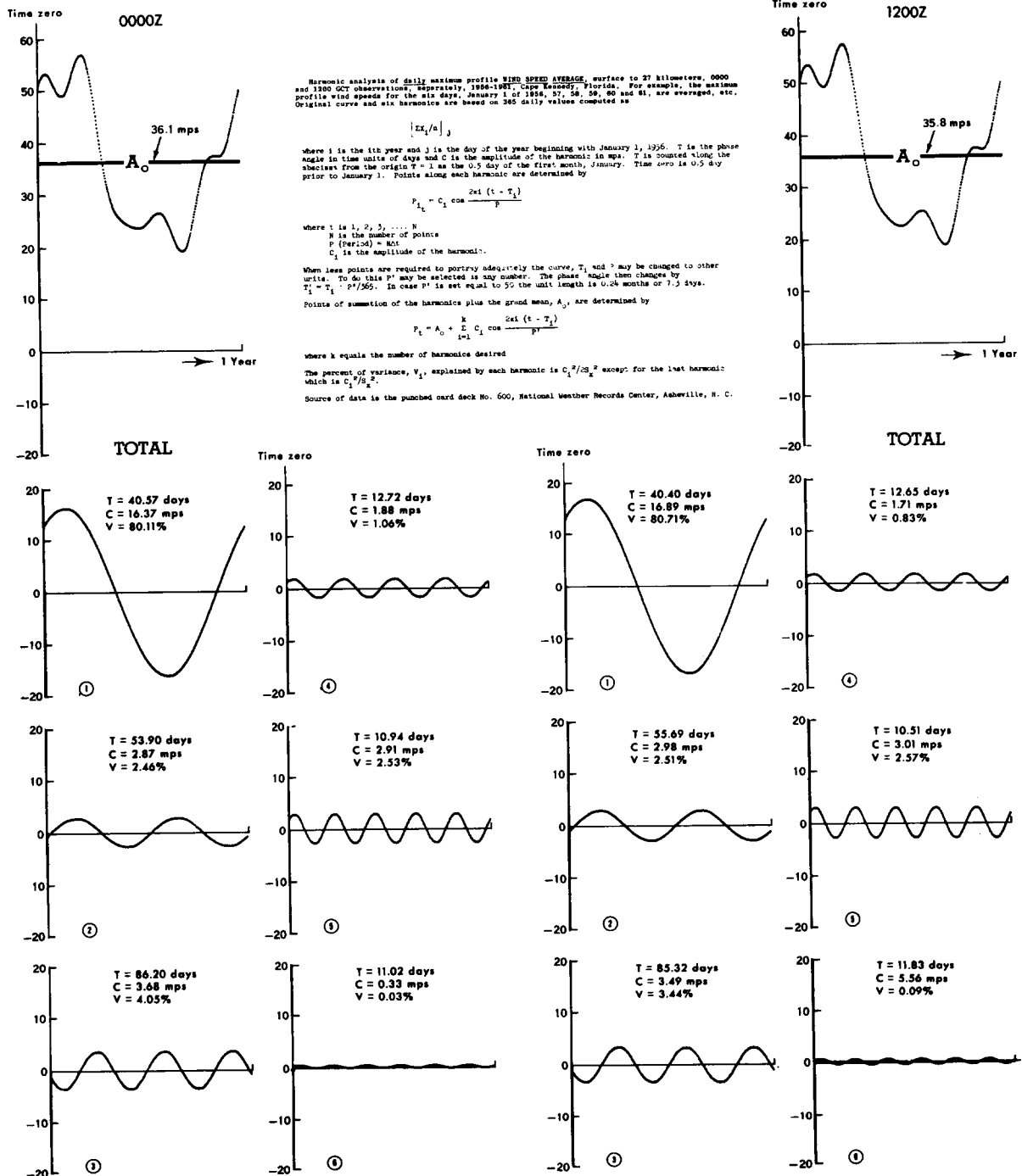
The input usually has been wind data from either Cape Kennedy, Florida or from the North American continent wind fields, but in one instance 500-mb heights and tropopause heights were examined as predictors. Use of other parameters is reserved for future investigation.

The first case above was restricted to the determination of daily means and variances from 6 years of data, or essentially a sample of six. These means and variances then were fitted by Fourier series (harmonic analysis). The insignificant harmonics were eliminated. Predictions for any date were made on the basis of the mean and variance for that date--computed from the harmonics. Tests for normality were made. Then the distributions were assumed to be distributed normally and independently for each date. The prediction is the mean plus or minus certain increments of the standard deviation in order to provide selected percentages (or quantiles) of the distribution. For decision purposes, then, a Monte Carlo process can be used to provide the predicted value.

Figures 1 and 2 show for Cape Kennedy, Florida, the harmonic analyses of the means and variances through the year for the 0000Z and the 1200Z observations. The shape of the corresponding harmonics implies that there is no difference between the 0000Z and the 1200Z sets of data. Figure 3 shows the summation of the significant harmonics

3

Fig. 1.

HARMONIC ANALYSIS OF DAILY PROFILE MAXIMUM WIND SPEEDS, SURFACE TO 27 KILOMETERS
CAPE KENNEDY, FLORIDA, 1956—1961

0000Z

Time zero

36.1 mps

A_o

1 Year

TOTAL

1200Z

Time zero

35.8 mps

A_o

1 Year

TOTAL

Harmonic analysis of daily maximum profile WIND SPEED AVERAGE, surface to 27 kilometers, 0000
and 1200 GCT observations, separately, 1956-1961, Cape Kennedy, Florida. For example, the maximum
profile wind speeds for the six days, January 1 of 1956, 57, 58, 59, 60 and 61, are averaged, etc.
Original curve and six harmonics are based on 365 daily values computed as

$$\left| \Sigma x_i / n \right|_j$$

where i is the ith year and j is the day of the year beginning with January 1, 1956. T is the phase
angle in time units of days and C is the amplitude of the harmonic in mps. T is counted along the
abscissa from the origin T = 1 as the 0.5 day of the first month, January. Time zero is 0.5 day
prior to January 1. Points along each harmonic are determined by

$$P_{i_t} = C_i \cos \frac{2\pi i (t - T_i)}{P}$$

where t is 1, 2, 3, ... N
  N is the number of points
  P (Period) = NΔt
  C_i is the amplitude of the harmonic.

When less points are required to portray adequately the curve, T_i and P may be changed to other
units. To do this P' may be selected as any number. The phase angle then changes by
$T_i' = T_i \cdot P'/365$. In case P' is set equal to 50 the unit length is 0.24 months or 7.3 days.

Points of summation of the harmonics plus the grand mean, A_o, are determined by

$$P_t = A_o + \sum_{i=1}^{k} C_i \cos \frac{2\pi i (t - T_i)}{P'}$$

where k equals the number of harmonics desired

The percent of variance, V_i, explained by each harmonic is $C_i^2/2S_x^2$ except for the last harmonic
which is C_i^2/S_x^2.

Source of data is the punched card deck No. 600, National Weather Records Center, Asheville, N. C.

0000Z harmonics

① T = 40.57 days, C = 16.37 mps, V = 80.11%

② T = 53.90 days, C = 2.87 mps, V = 2.46%

③ T = 86.20 days, C = 3.68 mps, V = 4.05%

④ T = 12.72 days, C = 1.88 mps, V = 1.06%

⑤ T = 10.94 days, C = 2.91 mps, V = 2.53%

⑥ T = 11.02 days, C = 0.33 mps, V = 0.03%

Time zero

1200Z harmonics

① T = 40.40 days, C = 16.89 mps, V = 80.71%

② T = 55.69 days, C = 2.98 mps, V = 2.51%

③ T = 85.32 days, C = 3.49 mps, V = 3.44%

④ T = 12.65 days, C = 1.71 mps, V = 0.83%

⑤ T = 10.51 days, C = 3.01 mps, V = 2.57%

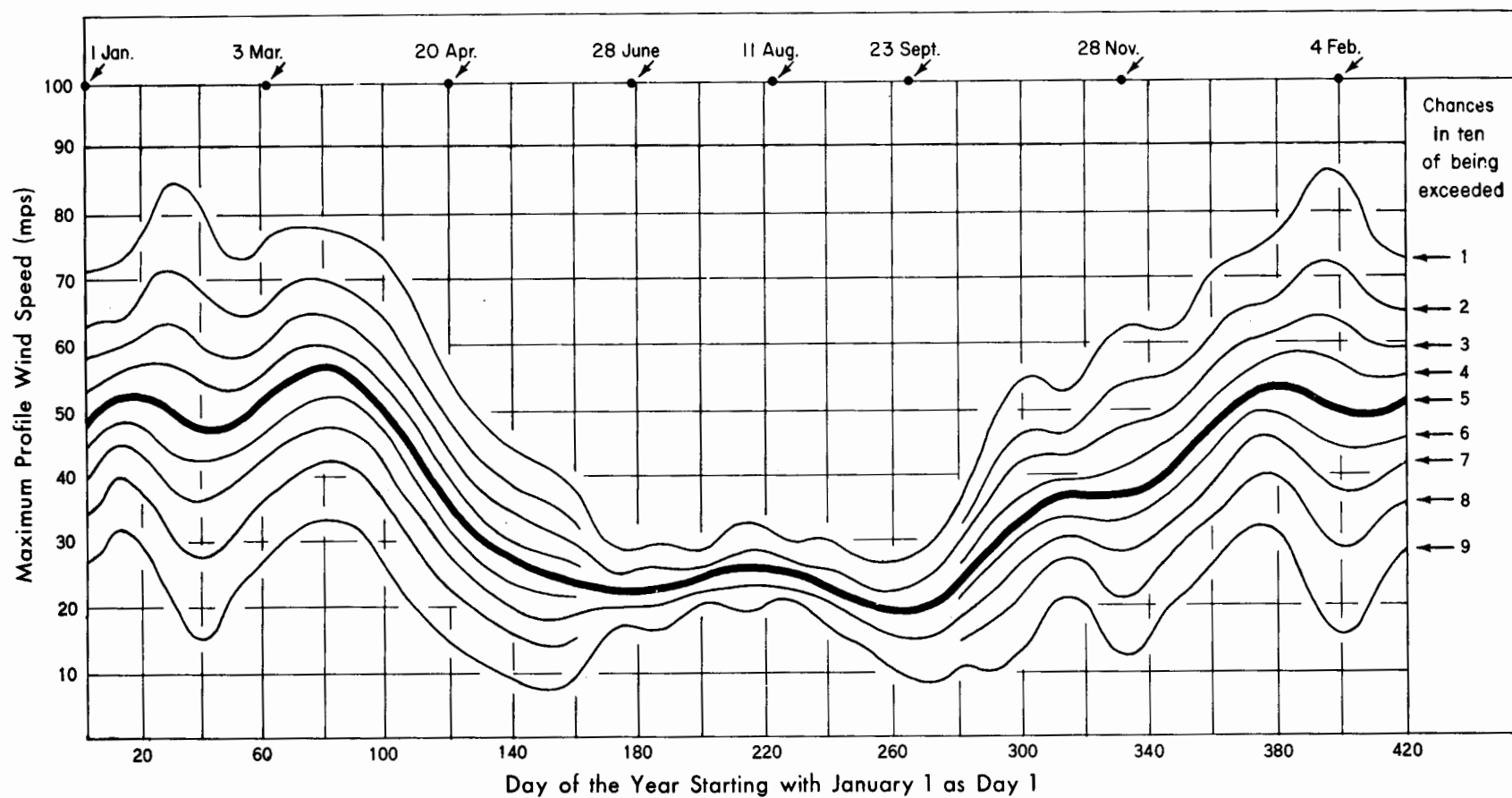⑥ T = 11.83 days, C = 5.56 mps, V = 0.09%

Time zero

Fig. 3. Cape Kennedy, Fla., Wind Profile Wind Maxima. Based on the period 1956-61, the chances in 10 are shown for a maximum wind speed of a profile to be exceeded.
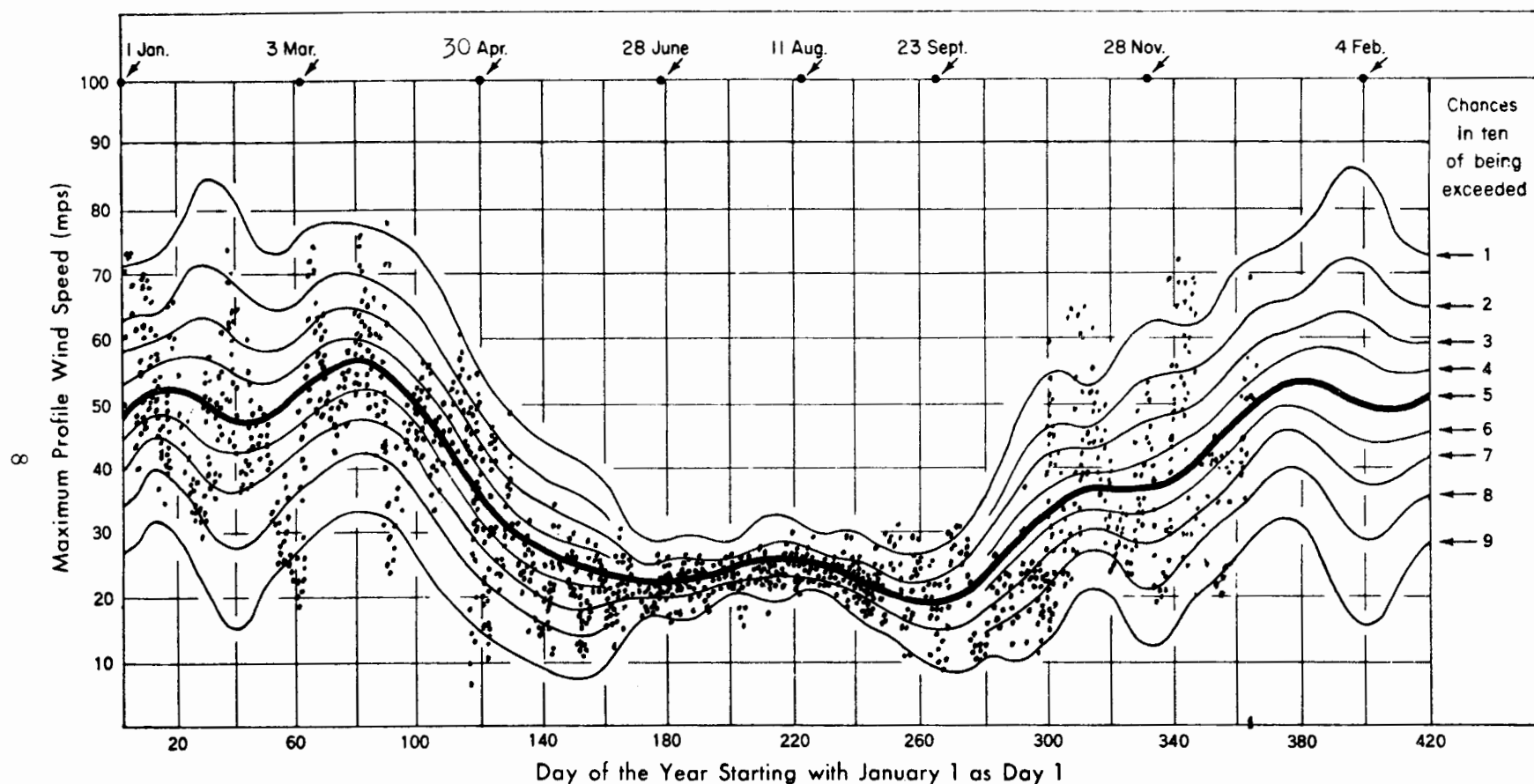
Fig. 4. Cape Kennedy, Fla., Wind Profile Wind Maxima. Based on the period 1956-61, the chances in 10 are shown for a maximum wind speed of a profile to be exceeded.

...) Maximum wind speed of the 0-27 km wind profile at Cape Kennedy, Florida. Based on 1962 data..

for the mean values and the variances. The central heavy line shows the seasonal march of the mean values. The deciles shown are derived from the respective variances. For each date there is an expected value, the mean, and for that same date the corresponding expected deciles have been computed and placed. Figures 4 and 5 have actual observed data for the respective years of 1962 and 1963 superposed on Figure 3. Examination of these figures shows that there appears to be a small periodicity in the wind. This feature will be discussed later.

In the second and third cases probabilistic envelopes of profiles from the surface to 100,000 feet were made. Individual profiles were described in terms of orthogonal polynomials. The orthogonal polynomial coefficients were obtained for each profile in a set of profiles. Multiple linear regression screening techniques were used with these coefficients to obtain predictive equations for coefficients of future profiles. These predicted orthogonal polynomial coefficients were used to construct or to synthesize profiles or envelopes of profiles.

The present study is restricted to the prediction of maximum winds in the maximum dynamic pressure region. This is the 10-15 km layer above the launching pads of the space vehicles. The process explored here is the Markov process.

Although prediction accuracy sought will be the best, the requirements ought to be made in terms of the potential available in the data. At first glance a requirement of a standard error of 1.5 mps or a range of plus or minus 5 mps does not seem too stringent. However, examination of the literature and unpublished works indicates that observational standard errors of 4 to 7 mps

7

are likely. Rapp [49] shows an observational standard error of about 2 mps when different telemetering systems are used from the same balloon train. Gabriel and Bellucci [23] show a standard error of about 3 mps in the layers above 300 mb. At altitudes of 7 to 15 km Anderson [2] indicates that for GMD-1A [54] equipment instrumental error is about 2.5 mps. A U. S. Navy report [53] implies that between 16 and 30 km a standard error of about 7 mps exists in the wind measurements. Plagge and Smith [47] indicate a standard error of about 3 mps at altitudes at or above 7 km.

Crutcher [11] shows that predictions of 300 mb winds at Omaha, Nebraska, and other points have a standard error of about 7 mps. Reed [50] shows forecast standard errors of 15 to 20 mps at higher altitudes. The U. S. Weather Bureau [55] at 8 locations over the U. S. at 7-8 km (25,000 feet) indicates predictive standard errors which average 7 mps.

Until newer and better instrumentation is available a standard error of measurement of 2 mps seems to be the best obtainable. This then provides an optimum minimum range of plus and minus 6 mps or a total range of 12 mps, though it was hoped that the forecasting procedures could be developed in these preliminary studies to reduce the standard error to at least 5 mps. This would be a reduction in variance of 10 to 1 (one order of magnitude) from that indicated by Reed [op. cit.]. Prediction to within 10 mps of the observed would be good. This would be a reduction in variance of an order of 2 to 1 (an order of magnitude of one-fifth).
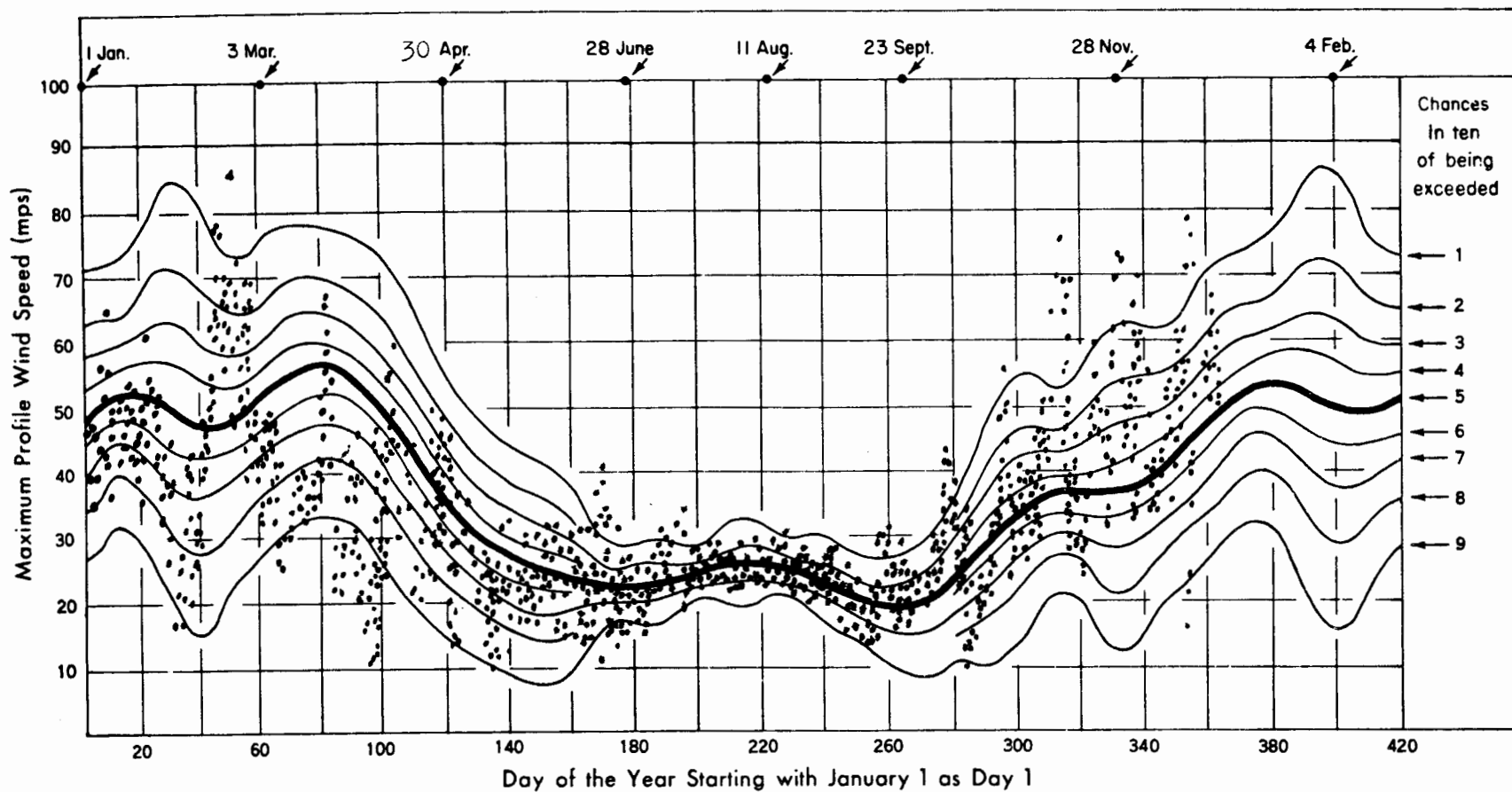
Fig. 5. Cape Kennedy, Fla., Wind Profile Wind Maxima. Based on the period 1956-61, the chances in 10 are shown for a maximum wind speed of a profile to be exceeded.

...) Maximum wind speed of the 0-27 km wind profile at Cape Kennedy, Florida. Based on 1963 data.

II. OBJECTIVE

Acceptable predictions of the maximum wind in the maximum
dynamic pressure region for space vehicles over Cape Kennedy are
to be sought for twelve-hour increments out to 120 hours through
the use of transition matrices of an operating Markov process or
Markov chain, M. P. or M. C. The predictions for this immediate
study will be made only from the Cape Kennedy upper wind data avail-
able from eight years of observations. They will be made within the
bounds of the noise imposed by instrumentation, observer error, and
of the operating thermodynamic systems. A prediction standard error
of 5 mps at this stage would be considered a success with GMD-1A
[op. cit.] equipment.

It is a well-established fact that persistence is a dominant
feature of weather systems. For a predictive scheme to be good and
useful it must be significantly better than either persistence or
climatology. Therefore, the results of some of the comparisons be-
tween the forecast models will be included in this paper.

III. DATA SOURCE

All meteorological data of the Environmental Science Services
Administration (ESSA), Air Force and Navy are stored at the National
Weather Records Center in Asheville, North Carolina. In addition to
the winds aloft data available in punched card decks there are decks
containing winds, temperature, moisture measurements and heights at
specified pressure levels. These data are referred to as thermodynamic
data. Reference manuals which describe these various sets of data are
available at ESSA-EDS-NWRC.

The development of prediction techniques requires sets of data

11

which are serially complete. The only deck or set of serially complete wind data is Card Deck 600 produced by ESSA-EDS-NWRC for NASA-MSFC-R-AERO-YT and NASA Langley for Cape Kennedy, Florida; Washington, D. C.; Norfolk, Virginia and Santa Monica, California. The periods vary from six years to ten.

The record for Cape Kennedy contains wind direction and speed in mps at 1 km levels from the surface through 27 km for the period 1956-1963. This can be considered to be a short term record that provides a sample of only eight. For a short period at the end of the eight years four observations per day are available. However, only two observations per day are used in this study. These are the 0000Z or 0300Z and the 1200Z or 1500Z observations.

IV. ENTROPY AND INFORMATION THEORY

Entropy and information theory play a vital role in the assessment of the potential of the transition matrices in a Markov process. It is pertinent to discuss these two concepts before discussion of the Markov process, though the discussion is necessarily brief.

A. Entropy

Getman and Daniels [25] interpreted the second law of thermodynamics to mean that a state of equilibrium is approached by all systems. In other words there is a tendency for all systems to approach a state of maximum probability. Entropy is a measure of the extent to which a system is random. It is a measure of disorder in a system; it is a measure of energy. Entropy increases when a system passes into a more random or less ordered state and conversely, it decreases when the system passes into a less random or more ordered state. This leads to the third law of thermodynamics, which says

that the entropy of a crystal at absolute zero is zero. The discussion above and that which now follows perhaps may be clearer when some analogies are made.

There is a need to have some measure, intuitive or otherwise, of the order and information in the meteorological systems under study. Consider a channel which may be used to communicate information from an input to an output signal. In general, some alphabet is assumed as indicated by Feinstein [18]. For perfect information and communication, a letter of the alphabet at the input would have a one-to-one correspondence to the output letter. It would not be necessary that there be a one-to-one correspondence between elements of the sets. The alphabets are such that a $\leftrightarrow$ b or A $\sim$ B are sets and an element "a" represents an element (letter) of set A {A} while "b" represents the corresponding element (letter) of set B {B}. An alphabetic letter at the output is associated with one and only one alphabetic letter at the input. If this occurs with each transmission, there is perfect communication. As there is complete order--no disorder--in this case, the entropy of the communication channel would be considered to be zero.

If there is some noise or disorder in the channel, then the entropy of the system would be something greater than zero. There would be some state at which the disorder in the system would be a maximum; the system would be completely random and usable (available) information would be nil.

Shannon [51] wrote the first paper on communications utilizing the concept of information theory. This paper is now a classic.

13

Shannon also employed the concept of entropy, though he and others after him call this "equivocation" rather than entropy.

A measure which satisfies the needs of this concept is given below for a single element; Shannon [op. cit.], Feinstein [op. cit.], Kullback [38], Masuyama [43], Baldwin [5].

$$H = -\sum_{i=1}^{c} p_i \ln p_i \qquad \text{IV.A.1}$$

where p_i is the probability of the system being in some one state, $\underline{i}$, of a number of states. The negative sign assures the positiveness of the quantity H, called equivocation or entropy. H will be called entropy throughout the remainder of this paper. If p_i is zero, $p_i \ln p_i$ is defined as zero. If there is only one state, the entropy is zero; that is, the system is fixed and there is no chance to pass outside the system. For example, if in the study of maximum winds in the maximum dynamic pressure region over Cape Kennedy, Florida, during January, the class interval chosen is 0-200 mps, and the maximum wind of all maxima recorded is 103 mps, then the likelihood estimate of the entropy

$$\hat{H} = -\sum_{i=1}^{c} (f_i/n) \ln (f_i/n) \qquad \text{IV.A.2}$$

will be zero. Here, f_i is the observed frequency in the state "i" and n is the total number of observations. In this example $\Sigma f_i = n$ and $c = 1$. The system is essentially in a fixed crystalline state. There is no additional information to be obtained from any observation either prior or post to the time of observation. Total information is available in the NOW condition. There is no need to

observe, to study, and to predict. The answer is known that the winds will be in the one state 0 to 200 mps. This then satisfies one boundary condition.

If all probabilities are equal in each and every state of two or more states, then we find the other bound, which is the maximum entropy, the maximum disorder, or the lack of any applicable information to permit forecasting into one of the categories. This maximum entropy is equal to the logarithm of the number of states. Thus, in a two-state matrix a measure of the maximum disorder or the maximum entropy would be the logarithm of two. In order to compare the entropies from one type of matrix to another, the entropies may be normalized by dividing each entropy by its respective possible maximum entropy. Thus, in a two-state matrix, the entropy would be divided by the logarithm of 2 while the entropy of a three-state matrix would be divided by the logarithm of 3.

The concept introduced by Shannon [op. cit.] and others seems to be satisfied intuitively by the entropy concept defined above.

Masuyama [op. cit.] discusses the likelihood estimate of entropy, its bias and its variance. He shows that the entropy behaves as a formal variance. This fact will be utilized later.

B. Information Theory

Karl Pearson [46] devised the now well-known and familiar χ^2 (Chi-squared) test to check the heterogeneity or goodness of fit of quantified classified data with respect to some theory specifying expected frequencies. Generally, the null hypothesis that there is no difference is tested.

Fisher [21], as a corollary development to the maximum likelihood criterion, proposed the log likelihood ratio criterion. Neyman and Pearson [44] developed this still further and showed that $-2 \ln \lambda$ was distributed as χ^2 with appropriate degrees of freedom where λ is the likelihood ratio criterion. Herdan [30] draws attention to the close relationship of χ^2 and H, the entropy. Woolf [60], Kupperman [42], Kullback et al [40] discuss this history in more detail.

Another statistic, "I," has been developed by Kullback and Liebler [41], used as $2\hat{I}$ by Kullback [op. cit.], discussed by Kupperman [op. cit.], and further developed by Kullback et al [40]. Kullback et al [40] called the $2\hat{I}$ a minimum discriminant information statistic (m.d.i.s.). Baldwin [4,5] independently developed the same statistic from a consideration of the channel capacitance in communication theory. He labeled his statistic "u", called it a dependence capacitance statistic (d.c.s.), and used it as $2n\bar{u}$.

If $p(x)$ is the multinomial distribution on a population of c classes, and $p'(x)$ is any other distribution on the population of c classes such that every possible observation from $p'(x)$ is also a possible observation from $p(x)$, then

$$I = u = \sum_{x_1 + \ldots + x_c = n} p'(x) \ln \frac{p'(x)}{p(x)} \qquad \text{IV.B.1}$$

Kullback et al [40] develop a maximum likelihood estimate $\hat{I}$ in terms of frequencies, and Baldwin [4,5] develops his estimate $\bar{u}$ in terms of probabilities. When this latter quantity is multiplied by the n-count of the sample,

$$\hat{I} = n\bar{u} = \Sigma_a f_a \ln \frac{f_a}{np_a} \qquad\qquad \text{IV.B.2}$$

where f_a is the frequency of observations in the a-th cell, np_a is

the theoretical or expected frequency of observations in the a-th

cell, and the summation is extended over all cells $a = 1,2,...$ The

expected probabilities $p_a > 0$, and their sum over all categories

$\Sigma_a p_a = 1$. For an $f_a = 0$ the quantity $0 \ln 0$ is defined as zero. The

$\hat{I} = n\bar{u}$ is found to be essentially equivalent to the $-\ln \lambda$ of Neyman

and Pearson [op. cit.].

By use of the approximation

$$\ln \frac{f_a}{np_a} \approx (1/2) \frac{f_a^2 - (np_a)^2}{f_a(np_a)}$$

Kullback et al [40] show that for p_a, $f_a > 0$

$$2\hat{I} = 2\Sigma_a f_a \ln \frac{f_a}{np_a} \approx \Sigma_a \frac{(f_a - np_a)^2}{np_a} \qquad\qquad \text{IV.B.3}$$

The last expression in IV.B.3 is the familiar χ^2 statistic. Thus,

$2\hat{I}$ and the equivalent statistics $2n\bar{u}$ and $-2\ln\lambda$ are distributed asymp-

totically as χ^2 with the appropriate number of degrees of freedom

as developed by Fisher [21].

Ku [36] discusses the inflated values of $2\hat{I}$ that occur with

the presence of zero cell frequencies. This inflation results from

the fact that for an $f_a = 0$, the m.d.i.s. (or d.c.s.) term on the

left side of the approximation in IV.B.3 is zero whereas the corres-

ponding χ^2 term on the right side is negative. It follows, therefore,

that $2\hat{I}$ is always greater than χ^2 when zero cell frequencies occur in

the sample. Ku [36] proposes an empirical correction factor of
subtracting one from the computed $2\hat{I}$ (or $2\bar{n\nu}$) for each zero cell
frequency in order to compensate for the inflated values. This
correction, however, is valid only if there are no more than a
few cells with zero frequencies [37].

The information statistics provide the measure of the transmission
capability of a communication channel. Perhaps the name selected by
Baldwin is a little more descriptive than that chosen by Kullback
et al [40]. The communication channel described by Feinstein [op.
cit.] has a certain capacitance or capability to transmit information.
The channel may have a certain noise level, yet the relationship
or history (or memory) which exists in the channel will control the
amount of usable information which is received as an output. The
word dependence capacity, therefore, implies the capacity and the
memory that is operating and which can be utilized. This dependence
capacity then is in part a measure of the Markovity of the system.
These entropy and information concepts are implicit throughout
all further discussions.

V. MARKOVITY AND PROBABILITY MATRICES

Transitional probabilities and Markov processes and chains are
discussed by Kolmogorov [33], Feller [19], Kullback [op. cit.], Chung
[10], Dynkin [17], Billingsley [7,8], Keeping [31], Wilks [58],
Kullback et al [40] and Bartlett [6]. Further bibliographic reference
may be made to Billingsley [7], who provides 113 references.

Application of such probabilities in the meteorological field
has been made by Andre [3], Gabriel and Neumann [24], Allen et al [1],
Caskey [9], Weiss [57], Feyerherm and Bark [20], Baldwin [4, 5],

18

Wiser [59], Crutcher and Orovitz [op. cit.], Quinlan [48], Grin-
gorten [29], and Godske [26, 27, 28]. The above references are
only a small portion of those available in the literature.

A. Some Preliminary Arguments

The Markov phenomenon in a continuous distribution is called a
process. In a discrete distribution it is called a chain. Weather,
in general, is a continuous persistent distribution, but because
measurements and time intervals essentially make it discrete, it can
be thought of as a Markov chain (M.C.).

A Markov process or chain (M.P. or M.C.) has a memory, history
influence, or persistent feature of m time periods. In many instances
and in many texts the concept of the Markov process or chain involves
no more than one time interval of history or memory. However, here
the system used is that of the m-th order. The system is called
an m-th order M.P. or M.C. In a first order M.C. with a time period
of one day, today's weather affects or controls tomorrow's weather
but not day-after-tomorrow's weather. Similarly, a second order M.C.,
if the time interval is one hour, implies that the weather an hour
ago and the weather now both influence or control the weather an
hour from now. The time interval is arbitrary and could be a week,
month, year, or decade. The problem generally will suggest the
interval(s) to be used.

A process is stationary if its distribution does not change even
though the sampled data may differ from sample to sample. Koopman
[34, 35] defines an M.P. or M.C. as stationary when the conditional
probabilities (probability of an event given a previous event) remain
fixed. He considers a series of successes or failures of an arbitrary

event under conditions such that the probability of success does not remain constant from trial to trial. Wadsworth et al [56] consider the theory of probability distributions in related trials. Their investigation is based on the works of Koopman [34, 35].

An M.P. or M.C. may be cyclic in that the process varies with some regular periodicity or return feature. In the realm of weather, processes through the year may be expected to show a periodic feature. There could be other periodic or aperiodic forcing functions modulating these which may be difficult to identify, much less to isolate and to remove. Even in the periodic case the required mathematics and arithmetic will be cumbersome. The principal periodicities considered in the data used in this investigation are:

1. diurnal

2. short term (weekly, monthly, ...)

3. seasonal

4. annual

In a study of ocean surface weather data [12] it is found that periodic functions are apparently existent in data collected over the years and which are treated as an annual ensemble. Following this, Baldwin [4] determined that the processes in the annual data are not stationary. Neither of these is unexpected. The usual method of avoiding the cyclic effects is to treat the data in monthly or shorter ensembles. This is followed even though partially truncated distributions may have been obtained. The null hypothesis of stationarity in the monthly data is not rejected and it is assumed that the process is indeed stationary. By an analogy, processes

through the day will be expected to be periodic but within an hourly interval the process may be essentially stationary.

No mathematics will be used in this report to isolate and remove periodic or aperiodic processes embedded in the data. The isolation and extraction of the periodic or aperiodic functions from the data is deferred to later research. It is assumed here that the processes involved are reasonably stationary over a month.

Ergodicity in the Markov system implies that it is possible to go from any state to any other state, yet that the process in time will converge to some determinable state. Parzen [45] indicates that if sample or time averages obtained from a record may be used as an approximation to the corresponding population average, then the process can be said to be ergodic. In general, weather processes converge to climatology. Any disruption in the convergence process in the weather will be damped in time and the process will continue its convergence to climatology. The disruptive influence disappears in time and subsides into the general trend of the total system. Thus, the climatological vector of a Markov distribution may be considered to be the ergodic vector. Ergodicity implies stationarity but the converse is not true.

B. Preparation of Probability Matrices

At this point a standard matrix format will help to visualize the procedures of forming an array. Table V.1 shows a matrix in terms of frequencies while Table V.2 shows a matrix in terms of empirical probabilities or relative frequencies. Table V.2 is the same as Table V.1 except that each frequency f_{ij} has been divided by n, the number of pairs of observations.

21

TABLE V.1 SCHEMATIC RAW FREQUENCY MATRIX

Table V.1 Schematic set-up of a two way, row and column, r x c,

contingency table in terms of frequencies. An f_{ij} is the frequency

of observation of the i-th row with the j-th column. $f_{i.}$ is the

frequency of the i conditions i.e., the summation of the i-th row

over all columns j. An $f_{.j}$ is the summation of frequencies of the

j-th column over rows i. The $f_{..}$ then is the summation of frequencies

over all rows and columns and is equivalent to n, the total number of

paired observations.

First criterion of classification NOW (i)	Second criterion of classification LATER (j)							
	1	2	3	...	j	...	c	Total
1	f_{11}	f_{12}	f_{13}	...	f_{1j}	...	f_{1c}	$f_{1.}$
2	f_{21}	f_{22}	f_{23}	...	f_{2j}	...	f_{2c}	$f_{2.}$
3	f_{31}	f_{32}	f_{33}	...	f_{3j}	...	f_{3c}	$f_{3.}$
...	...	...	...	...	...	...	...	...
i	f_{i1}	f_{i2}	f_{i3}	...	f_{ij}	...	f_{ic}	$f_{i.}$
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
r	f_{r1}	f_{r2}	f_{r3}	...	f_{rj}	...	f_{rc}	$f_{r.}$
Total	$f_{.1}$	$f_{.2}$	$f_{.3}$	...	$f_{.j}$	...	$f_{.c}$	$f_{..} = n$

TABLE V.2   SCHEMATIC RELATIVE FREQUENCY MATRIX

Table V.2 Schematic set-up of a two way, row and column, r x c, contingency table in terms of empirical probabilities or relative frequencies where $p_{ij} = f_{ij}/f_{..} = f_{ij}/n$, n or $f_{..}$ is the number of paired observations, and f_{ij} is the frequency of observation of the i-th row/j-th column combination.

First criterion of classification NOW (i)	Second criterion of classification LATER (j)							Total
	1	2	3	...	j	...	c	
1	p_{11}	p_{12}	p_{13}	...	p_{1j}	...	p_{1c}	$p_{1.}$
2	p_{21}	p_{22}	p_{23}	...	p_{2j}	...	p_{2c}	$p_{2.}$
3	p_{31}	p_{32}	p_{33}	...	p_{3j}	...	p_{3c}	$p_{3.}$
...	...	...	...	...	...	...	...	...
i	p_{i1}	p_{i2}	p_{i3}	...	p_{ij}	...	p_{ic}	$p_{i.}$
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
r	p_{r1}	p_{r2}	p_{r3}	...	p_{rj}	...	p_{rc}	$p_{r.}$
Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	...	$p_{.j}$	...	$p_{.c}$	1.00

A schematic of a transitional probability matrix is shown in Table V.3. A transitional probability is a probability for the occurrence of a later event given that an initial condition has occurred. The matrix is prepared easily by dividing the f_{ij} by the respective row frequencies $f_{i.}$ of Table V.1. It should be noted that the sum of the transitional probabilities over a row $p_{i./i}$ must equal 1.00.

The matrix shown in Table V.3 is used as the basic scheme for predicting categories of maximum winds. Only a category, and not a specific wind speed, can be forecast. The requirement of establishing suitable classes essentially makes the continuous wind distribution discrete.

For ease of reference, following Baldwin [4], the transitional probability matrices prepared for this study may be identified as $_\beta[p]_t$ where

1. β indicates the period covered, such as $\underline{s}$ for 12-hour, $\underline{d}$ for 24-hour, $\underline{w}$ for week, $\underline{m}$ for month, $\underline{a}$ for annual.

2. [p] is the transitional probability matrix.

3. $\underline{t}$ is the time interval of the matrix such as 12 for 12 hours, 24 for 24 hours, 36 for 36 hours, etc.

C. Tests of Transitional Probability Matrices

1. Stationarity test

If $d_1 \ldots d_t \ldots$ is a stationary M.P. or M.C. with a transition matrix [p] and an absolute, climatological, marginal, or ergodic vector [π], then Baldwin [4] shows that

$$[\pi] \, [p] = [\pi] \qquad\qquad\qquad \text{V.C.1}$$

24

TABLE V.3 SCHEMATIC PREDICTION MATRIX

Table V.3 Schematic set-up of a two way, row and column, r x c, contingency table in terms of empirical transitional probabilities $p_{ij/i} = f_{ij}/f_i$ where f_{ij} is the frequency of observation of the i-th row/j-th column combination, and f_i is the summation of the i-th row frequencies over all columns j.

First criterion of classi- fication NOW (i)	Second criterion of classification LATER (j)							
	1	2	3	...	j	...	c	Total
1	$p_{11/1}$	$p_{12/1}$	$p_{13/1}$	...	$p_{1j/1}$	...	$p_{1c/1}$	$p_{1./1}$
2	$p_{21/2}$	$p_{22/2}$	$p_{23/2}$	...	$p_{2j/2}$	...	$p_{2c/2}$	$p_{2./2}$
3	$p_{31/3}$	$p_{32/3}$	$p_{33/3}$	...	$p_{3j/3}$	...	$p_{3c/3}$	$p_{3./3}$
...	...	...	...	...	...	...	...	...
i	$p_{i1/i}$	$p_{i2/i}$	$p_{i3/i}$	...	$p_{ij/i}$	...	$p_{ic/i}$	$p_{i./i}$
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
r	$p_{r1/r}$	$p_{r2/r}$	$p_{r3/r}$	...	$p_{rj/r}$	...	$p_{rc/r}$	$p_{r./r}$

Thus, the quantity $[\pi]$ $[p]$ - $[\pi]$ can be used to test the stationarity of a process. If the quantity for a given process is zero, or at least not significantly different from zero, then the process can be assumed to be stationary. Not enough experience has been gained to establish a firm estimate as to how large a difference from zero can be accepted before a decision is made that the operating system is not stationary. Obviously, the number is a function of the number of observations as well as of the number of categories.

2. Distribution tests

The information statistics can be used to test whether or not a given distribution is statistically the same as a specified theoretical distribution. From IV.B.3 and the discussion following it, the test, under the null hypothesis that the expected and observed cell frequencies are the same, is

$$2\hat{I} = 2n\bar{u} = 2\sum_a f_a \ln \frac{f_a}{np_a} - b \qquad \text{V.C.2}$$

where f_a is the observed frequency in the $\underline{a}$-th cell, n is the number of observations in the sample, p_a is the theoretical probability of the $\underline{a}$-th cell, and b is the number of cells with zero frequencies. The test statistic is asymptotically distributed as χ^2. Since the total number of degrees of freedom of the system ν_t is a - 1, and the number of restrictions placed upon the system is ν_{np_a} (the number of degrees of freedom associated with np_a), the appropriate number of degrees of freedom ν with which to enter the χ^2 table is

$$\nu = \nu_t - \nu_{np_a} = a - 1 - \nu_{np_a} \qquad \text{V.C.3}$$

26

The subtle point of rejection or of non-rejection of hypotheses arises. The non-rejection of a hypothesis does not mean that the hypothesis is accepted in totality or finality. A stronger or more precise test may indicate that the hypothesis should be rejected. The precision of the m.d.i.s. or d.c.s. test statistic appears to be dependent upon the n-count of the sample. If a $2\hat{I}$ based on $\sum_a (f_a > 0) = n$ sample observations is y, then the $2\hat{I}$ based on kn sample observations is ky, where k is a constant multiplier of each frequency of the original distribution. For a constant empirical probability distribution the $2\hat{I}$ is thus a monotonic increasing function of the sample size n. The degrees of freedom, however, are independent of n and therefore do not change as $2\hat{I}$ varies. Further research is needed to develop a quantitative measure of the precision of the test.

One of the uses of the test statistic is that of comparing an observed frequency distribution to an equiprobable distribution. That is, it is compared with the maximum entropy possible in the system. Under the null hypothesis the probability of being in any cell is the same as that of being in any other cell. Setting p_a equal to a constant $1/g$, where g is the number of cells $a = 1, 2, \ldots, g$, the proper test statistic is

$$2\hat{I} = 2n\bar{u} = 2\sum_a f_a \ln \frac{gf_a}{n} - b, \quad \nu = g - 1 \qquad \text{V.C.4}$$

which breaks down into the components

$$2\hat{I} = 2n\bar{u} = 2\sum_a f_a \ln f_a + 2n \ln g - 2n \ln n - b \qquad \text{V.C.5}$$

These equations can be computed best by the use of electronic computers. However, table look-up permits advantageous use of desk calculators.

Kullback [op. cit.] and Masuyama [op. cit.] provide tables of ln n,

n ln n, and $n(\ln n)^2$, up to n=1000. Woolf [op. cit.] provides tables of

2n ln n up to n=2009 while Kullback et al [39]provide tables of 2n ln n

up to n=10,000.

A two-dimensional distribution can be expressed by a matrix of rows

i = 1,2, ..., r and columns j = 1,2, ..., c where the rows represent the

first dimension or criterion of classification, and the columns represent

the second criterion. If the two criteria are independent, then p_a =

$p_{ij} = p_i p_j$. Under the null hypothesis of independence between rows and

columns,

$$\hat{2I} = 2n\bar{u} = 2\sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} \ln\frac{f_{ij}}{np_i p_j} - b, \quad \nu = (r-1)(c-1) \quad V.C.6$$

The p_i and p_j can be estimated by the relative frequencies of the data

sample $f_{i.}/n$ and $f_{.j}/n$, respectively. The notation used is that of

Table V.1, V.2 and V.3. The statistic then becomes

$$\hat{2I} = 2n\bar{u} = 2\sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} \ln\frac{nf_{ij}}{f_{i.} f_{.j}} - b \qquad V.C.7$$

In a three-way classification, or cubical matrix with r rows, c

columns, and d depths, the d.c.s. for testing independence among the

three criteria is

$$\hat{2I} = 2n\bar{u} = 2\sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{d} f_{ijk} \ln\frac{n^2 f_{ijk}}{f_{i..} f_{.j.} f_{..k}} - b, \quad \nu = rcd-r-c-d + 2 \quad V.C.8$$

Extensions to higher way classifications are made easily, but the study

of the marginal distributions becomes rather complex or involved.

Such studies are the analysis-of-information studies proposed by Kull-

back et al [40] and are analagous to the analysis of variance with

an acronym ANOVA. A corresponding acronym could be ANOINF.

A sequence of observations can be tested against the hypothesis

that it is a stationary Markov chain of specified order m. Successive pairs ij of observations of the occurrences of the states or class intervals of a M.C. can be distributed in a two-way contingency table such that the first state of the pair is the row category i or the NOW condition, and the second state of the pair is the column category j or the LATER condition. The tally of overlapping pairs of observations can be represented as in Table V.1. If overlapping triplets are considered where an f_{ijk} is the frequency of observations of a PREVIOUS state i, NOW state j, and LATER state k, then a cubical or three dimensional matrix can be constructed. The process can be extended out to any dimension 1, 2, 3, ..., and all dimensions will have necessarily the same number of states or class intervals.

A stationary M.C. is determined completely by an initial probability vector and a matrix of transition probabilities. If a sequence of three observations i, j, and k is a representation of a M.C. of order one, then

$$p_{ijk} = (p_i)(\frac{p_{ij}}{p_i})(\frac{p_{jk}}{p_j}) \qquad\qquad V.C.9$$

where p_{ijk} is the probability of the whole sequence under the first order assumption, p_i is the initial probability or the probability of the first observation i, p_{ij}/p_i is the transition from i to j, and p_{jk}/p_j is the transition from j to k. Since a longer initial sequence would require more transitions, V.C.9 generalizes to

$$p_{ijk...yz...} = (p_i)(\frac{p_{ij}}{p_i})(\frac{p_{jk}}{p_j}) ...(\frac{p_{yz}}{p_y})... \qquad\qquad V.C.10$$

29

If the theoretical probabilities are not known, empirical probabilities should be used. The format of the matrix of empirical transition probabilities is shown in Table V.3.

Under a second order assumption a third observation is dependent upon the previous two observations. A transition probability can be represented by p_{ijk}/p_{ij}. The initial probability used is that for the first two observations; therefore,

$$p_{ijkl...xyz...} = (p_{ij})(\frac{p_{ijk}}{p_{ij}})(\frac{p_{jkl}}{p_{jk}}) \cdots (\frac{p_{xyz}}{p_{xy}}) \cdots \qquad V.C.11$$

Extension of the transition theory to higher orders is easily made.

Substituting a $p_{ijk...}$ for p_a enables V.C.2 to test whether an observed frequency distribution is a realization of a stationary M.C. of order specified by the $p_{ijk...}$. The null hypothesis is that the observed distribution is of order m within the assumption that it is of order m + 1. In other words a M.C. of order two, for example, has probabilities

$$p_{ijk} = (p_{ij})(\frac{p_{ijk}}{p_{ij}}) \qquad V.C.12$$

but the null hypothesis of order one implies that

$$\frac{p_{ijk}}{p_{ij}} = \frac{p_{jk}}{p_j} \qquad V.C.13$$

It follows, therefore, that

$$p_{ijk} = (p_{ij})(\frac{p_{jk}}{p_j}) = (p_i)(\frac{p_{ij}}{p_i})(\frac{p_{jk}}{p_j}) \qquad V.C.14$$

A given stationary sequence of observations is tested for 0, 1, 2, ... order. If the computed $2\hat{I}$ or $2m\bar{u}$ are ordinate values plotted against

monotonic increasing orders on the abscissa, then the order m of the
process operating in the sequence is that value of the abscissa for
which the ordinate value of the curve first falls below the chosen
χ^2 rejection level. Any order $<m$ will not utilize all the information
that can be gleaned from the "history" of the data. On the other hand
the additional "history" incorporated into any order $>m$ will not
lower the entropy or increase the information that is inherent in the
m-th order process.

When relative frequencies are used as estimates of the theoretical
initial and transition probabilities p_{ijk}, the m.d.i.s. and the d.c.s.
diverge from equality. The difference results from the assumptions
made about the behavior of the empirical initial probabilities.
Baldwin [5] places the restriction of independence upon these proba-
bilities such that

$$\frac{f_{ij}}{n} = (\frac{f_i}{n})(\frac{f_j}{n}), \quad \frac{f_{ijk}}{n} = (\frac{f_i}{n})(\frac{f_j}{n})(\frac{f_k}{n}), \quad \ldots \qquad \text{V.C.15}$$

Kullback et al [40] provide a more general test in that they do not
restrict the initial probabilities. The degrees of freedom with
which to enter the χ^2 tables depend upon the test statistic used.
For the m.d.i.s. or $2\hat{I}$

$$\nu = s^m(s^t - ts + t - 1) \qquad \text{V.C.16}$$

and for the more restrictive d.c.s. or $2n\bar{u}$

$$\nu = s^m(s^t - ts + t) - m(s - 1) - 1 \qquad \text{V.C.17}$$

where s is the number of states or class intervals into which each
observation can fall, m is the order of Markovity against which the
observed frequency matrix or sequence is being tested, and t is the

31

order m subtracted from the total length of the sequence i, j, k,

The minimum length of the total sequence must be m + 2.

3. Comparison between matrices

Baldwin [4] indicates some of several arrays which may be
used to determine whether the differences between matrices could
be considered to be significant. Let $P = (p_{ij})$ and $Q = (q_{ij})$ be
any two n x n matrices. The following quantities are measures of the
closeness to equality of P and Q.

a. <u>The Euclidean Distance</u>

$$|| P - Q ||_2 = \left[\sum_i \sum_j (p_{ij} - q_{ij})^2 \right]^{1/2} \qquad \text{V.C.18}$$

b. <u>The Norm Distance</u>

$$|P,Q| = \max_{(i,j)} |p_{ij} - q_{ij}| \qquad \text{V.C.19}$$

c. <u>The Mean Difference</u>

$$\overline{(P,Q)} = \frac{1}{n^2} \sum_i \sum_j |p_{ij} - q_{ij}| \qquad \text{V.C.20}$$

An important quantity associated with the mean difference is
the

d. <u>Variance of the Differences</u>

$$\sigma^2 (P,Q) = \frac{1}{n^2} \sum_i \sum_j \left[|p_{ij} - q_{ij}| - \overline{(P,Q)} \right]^2 \qquad \text{V.C.21}$$

Still another metric set would be the comparison of entropy and
the variance of the respective matrices. This is only suggested here
and will be investigated later. An estimate of an entropy, however,
would be, after Masuyama [op. cit.],

32

$$\hat{H} = - \sum_{i=1}^{k} (f_i/n) \ln (f_i/n) \qquad\qquad \text{V.C.22}$$

with

$$\hat{\sigma}_{\hat{H}}^2 = \left\{ \sum_{i=1}^{k} (f_i/n) [\ln (f_i/n)]^2 - \hat{H}^2 \right\} /n \qquad \text{V.C.23}$$

As indicated previously, Masuyama [op. cit.] states that the

entropy likelihood estimate is similar to a formal variance. It will

be necessary to check the behavior of the entropies for these are

bounded at zero by definition. The normalized entropies are bounded

by zero and one by definition. Perhaps a logarithmic transformation

of the entropies or an arc sine transformation of the normalized

entropies would be in order. From the above similarity the variance

ratio Z-test of Fisher [22], renamed the F-test by Snedecor [52],

would suffice. The above suggestions are for future investigation

for they may have been reported on already and the authors are unaware

of such reports.

VI. COMPUTATIONS

The maximum winds taken at 12-hour intervals in the 10-15 km

layer over Cape Kennedy, Florida, during the eight-year period

1956-1963 are used to develop transition probability matrices.

These matrices are of various sizes and dimensions.

A. Two-state Matrices

Two-state manually compiled matrices of dimension r = 2, 3, 4, 5

are studied primarily in order to explore the nature of the information

statistics and to verify some of the properties of these statistics.

Only January data are used. In one set of arrays the data are

dichotomized into states of winds <70 mps and winds ≥ 70 mps,

33

and in another set the two classes consist of winds <43 mps and
winds ≥ 43 mps.

The two-dimensional matrices have little predictive value
because the only possible forecasts are those of persistence and
of climatology. This can best be seen in **Table Vl.1**, which shows
the only four possible forecast schemes. In part a the process
operating is persistent in that the initial state will always be
forecast. A "negative" persistence is indicated in part b since
$j = 2$ for $i = 1$ and $j = 1$ for $i = 2$. Parts c and d show the two
climatological forecasts. In these cases the value of j is inde-
pendent of the initial state i.

An r-dimensional matrix with the same s states in each dimension
represents a Markov chain of order $m = r - 1$. Kemeny and Snell [32]
show that an m-th order chain can be reduced to a first order chain
and represented by a two dimensional matrix with s^m states. The
expansion process can be illustrated as follows. Consider a two-state
$(s = 2)$ M.C. of second order $(m = 2)$ that is depicted by an array
with $r = 3 = (\underline{i},\underline{j},\underline{k})$. The states corresponding to each dimension
are s_i, s_j and s_k. In the expanded chain $m = 1$ and $r = 2$ with the
first dimension containing the pair of states $(s_i, s_j) = s_{ij}$ and
the second dimension containing the pair of states $(s_j, s_k) = s_{jk}$.
Thus, the original two-state second order $(\underline{i} \times \underline{j} \times \underline{k})$ matrix is
transformed into a four-state first order $(\underline{ij} \times \underline{jk})$ matrix. The
reduction of a M.C. to first order facilitates the retrieval of useful
information from the system at the expense of greatly increasing the
complexity of the computations.

TABLE V1.1 Two-state, two dimensional transitional probability matrices with rows representing the NOW states i = 1, 2 and the columns representing the LATER states j = 1, 2. For a given i the forecast value of j is underlined. The forecast is that value of j given i which has the highest probability occurrence.

		LATER state j				LATER state j	
		1	2			1	2
NOW	1	<u>.70</u>	.30	NOW	1	.30	<u>.70</u>
state				state			
i	2	.30	<u>.70</u>	i	2	<u>.70</u>	.30

a) Persistence   b) "Negative" Persistence

		LATER state j				LATER state j	
		1	2			1	2
NOW	1	<u>.70</u>	.30	NOW	1	.30	<u>.70</u>
state				state			
i	2	<u>.70</u>	.30	i	2	.30	<u>.70</u>

c) Climatology   d) Climatology

B. Three-state Matrices

Manually tabulated $r = 2, 3, 4, 5$ dimensional matrices with

wind states $0 \leq s_1 < 43$ mps, 43 mps $\leq s_2 < 70$ mps, 70 mps $\leq s_3$ are

tested for stationarity and order of Markovity $m = 0, 1, 2, 3$. Only

January data are examined.

The test statistic computed is the $2\hat{I}$ of V.C.2. The p_a takes

on the values of the empirical transitional and less restrictive

initial probabilities that are appropriate for the order being tested.

When the expected probabilities are greater than zero for all cells

of a matrix, the degrees of freedom are evaluated from V.C.16. If

some of these probabilities are zero, however, V.C.16 is invalid and

the degrees of freedom ν are found by considering the degrees of

freedom associated with each independent component of the logarithmic

argument of V.C.2.

In order to check the results of m.d.i.s. order test of Kullback

et al [40] the normalized entropies R of the r-dimensional matrices

are computed and plotted on a graph of R vs r. The value of r for

which R first becomes nearly constant as r increases should be equal

to $m + 2$, where m is the order of the operating system.

C. 6-, 11-, 22-state Matrices

The maximum winds are separated into 6, 11 and 22 classes or

states (0-20 mps, 21-40 mps, ..., 101-120 mps), (0-10 mps, 11-20 mps,

..., 101-110 mps) and (0-5 mps, 6-10 mps, ..., 106-110 mps), respectively.

These data are processed by electronic computers to provide period-

month lag matrices of NOW against LATER observed conditions in time

intervals of 12 hours out to a maximum of 120 hours or lag $t = 10$

time intervals. The arrays contain frequencies of occurrence,

relative frequencies and empirical transition probabilities for each cell. The marginal or climatological vectors are also included.

For each matrix the following quantities are computed:

1. $[\pi]$ $[p]$ - $[\pi]$ to test for stationarity

2. entropy H and normalized entropy R

3. $2n\bar{u}$ for independence between NOW and LATER conditions (not corrected for zero frequencies)

The last quantity is used in the method suggested by Baldwin [5] for approximating the Markovity of a system. The $2n\bar{u}$ for each of the ten s-state lag matrices for a period-month is plotted as an ordinate value against the lag or increasing time interval on the abscissa. A horizontal line depicting the χ^2 rejection level is drawn. A point above the line indicates dependence or correlation between the NOW and lag t observations. Similarly, a point below the line indicates independence between the NOW and lag t observations. A smooth curve drawn through the points will cross the rejection line once if there is no periodicity in the data. If this intersection is at lag t, then all observations at a lag greater than t will not be influenced by the NOW condition. The Markovity of a system therefore is estimated to be the value of the lag at which the $2n\bar{u}$ curve intersects the χ^2 rejection curve. The method is only an approximation because the interactions or partial correlations between lags are not considered. It is probable in a stationary process that the information indicated by a significant value at a large lag will be extracted and provided for by a lower order process. The $2n\bar{u}$ versus lag graphs for the Cape Kennedy, Florida, period-month maximum wind data are held in file at ESSA-EDS-NWRC.

Kemeny and Snell [op. cit.] show that under the assumption of a first order stationary process [p] the two-step transition matrix of the process is given by $[p]^2$, the three-step transition matrix is given by $[p]^3$ ... In other words the partial correlations are zero and

$$_{\beta}[p]_t = \left(_{\beta}[p]_1\right)^t \qquad\qquad \text{VI.C.1}$$

where $_{\beta}[p]_1$ is the lag one correlation matrix and t is the transition step. For example, if $_{\beta}[p]_1$ is a 12-hour transition matrix, then $\left(_{\beta}[p]_1\right)^3$ is the 36-hour transition matrix. Wadsworth et al [op. cit.] discuss a phase of this procedure that is based on Koopman's work [34].

In order to test the validity of a first order assumption of 12- and 24-hour transition matrices, $_{\beta}[p]_1$ and $_{\beta}[p]_2$ are raised to powers and then are compared to the corresponding observed conditional probabilities. No products of two different interval matrices are compared to the corresponding observed conditional probabilities; that is, the product of the 12-hour and the 24-hour matrices are not compared to the observed 36-hour transitional probability matrix. The quantities $\left(_{\beta}[p]_1\right)^t - \left(_{\beta}[p]_t\right)^1$ and $\left(_{\beta}[p]_2\right)^t - \left(_{\beta}[p]_{2t}\right)^1$ are computed to test the model. If the differences are not significantly different from zero, then the first order assumption is valid.

The above computations are tabulated and held in file both at ESSA-EDS-NWRC and at NASA-MSFC-R-AERO-YT. The results of a few of the calculations have been extracted from the tabulations as examples and are shown in Table VI.2.a-e.

D. m-th Order Six-state Matrices

The maximum winds over Cape Kennedy, Florida, are separated into

TABLE VI.2.a   January transition probability matrices for the maximum winds in the 10-15 km

layer over Cape Kennedy, Florida, 1956-1963. The speed class intervals are 0-20 mps, 21-40 mps,

..., 101-120 mps. The intervals are numbered 1 through 6. The prior (NOW) conditions are listed

vertically with the post (LATER) conditions listed horizontally. The matrices are identified as

$_\beta[p]_t$ where β is the month and t is the lag in time intervals of 12 hours. $\beta = 1$ (January);

t = 1, 2, ..., 10. The climatological or marginal vector is $[\pi]$. The product of the marginal

vector and the transition matrix is shown as $[\pi][p]$. The difference $[\pi][p] - [\pi]$ is then shown.

$_1[p]_1$ LATER

$_1[p]_2$ LATER

		1	2	3	4	5	6	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.000	.745	.246	.009	.000	.000	.000	.666	.281	.053	.000	.000
NOW	3	.000	.118	.764	.118	.000	.000	.000	.158	.699	.132	.011	.000
	4	.000	.000	.327	.622	.051	.000	.000	.020	.418	.491	.071	.000
	5	.000	.000	.000	.364	.545	.091	.000	.000	.182	.545	.182	.091
	6	.000	.000	1.000	.000	.000	.000	.000	.000	.000	1.000	.000	.000
$[\pi]$		.000	.236	.542	.198	.022	.002	.000	.244	.534	.196	.024	.002
$[\pi][p]$		.000	.240	.539	.197	.022	.002	.000	.251	.528	.195	.024	.002
$[\pi][p] - [\pi]$		.000	.004	-.003	-.001	.000	.000	.000	.007	-.006	-.001	.000	.000

$_1[p]_3$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.605	.316	.079	.000	.000
NOW 3	.000	.188	.661	.140	.011	.000
4	.000	.031	.459	.418	.082	.010
5	.000	.000	.182	.727	.091	.000
6	.000	.000	.000	.000	1.000	.000
$[\pi]$	.000	.248	.530	.194	.026	.002
$[\pi][p]$	.000	.256	.522	.194	.026	.002
$[\pi][p] - [\pi]$	.000	.008	-.008	.000	.000	.000

$_1[p]_4$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.553	.368	.079	.000	.000
NOW 3	.000	.206	.617	.162	.015	.000
4	.000	.051	.520	.337	.082	.010
5	.000	.000	.091	.818	.091	.000
6	.000	.000	.000	.000	1.000	.000
$[\pi]$	.000	.250	.528	.192	.028	.002
$[\pi][p]$	.000	.257	.520	.193	.028	.002
$[\pi][p] - [\pi]$	.000	.007	-.008	.001	.000	.000

$_1[p]_5$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.561	.360	.079	.000	.000
NOW 3	.004	.199	.599	.176	.022	.000
4	.000	.061	.531	.337	.061	.010
5	.000	.000	.273	.545	.182	.000
6	.000	.000	.000	.000	1.000	.000
$[\pi]$	.002	.250	.522	.194	.030	.002
$[\pi][p]$	.002	.256	.514	.193	.031	.002
$[\pi][p] - [\pi]$	.000	.006	-.008	-.001	.001	.000

$_1[p]_6$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.009	.508	.404	.079	.000	.000
NOW 3	.004	.213	.574	.191	.018	.000
4	.000	.082	.510	.316	.082	.010
5	.000	.000	.364	.454	.182	.000
6	.000	.000	.000	.000	1.000	.000
$[\pi]$	.004	.250	.516	.196	.032	.002
$[\pi][p]$	.004	.253	.509	.195	.033	.002
$[\pi][p] - [\pi]$	.000	.003	-.007	-.001	.001	.000

$_1[p]_7$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.026	.483	.412	.079	.000	.000
3	.000	.221	.562	.191	.026	.000
4	.000	.102	.511	.306	.071	.010
5	.000	.000	.364	.454	.182	.000
6	.000	.000	.000	.000	1.000	.000
[π]	.006	.252	.512	.194	.034	.002
[π][p]	.007	.255	.503	.193	.035	.002
[π][p] - [π]	.001	.003	-.009	-.001	.001	.000

$_1[p]_8$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.035	.482	.430	.053	.000	.000
3	.000	.221	.544	.206	.029	.000
4	.000	.092	.531	.316	.051	.010
5	.000	.091	.454	.182	.273	.000
6	.000	.000	.000	.000	1.000	.000
[π]	.008	.252	.512	.192	.034	.002
[π][p]	.009	.255	.504	.186	.036	.002
[π][p] - [π]	.001	.003	-.008	-.006	.002	.000

$_1[p]_9$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.044	.491	.404	.061	.000	.000
3	.000	.217	.566	.191	.026	.000
4	.000	.082	.531	.316	.061	.010
5	.000	.000	.545	.091	.364	.000
6	.000	.000	.000	1.000	.000	.000
[π]	.010	.248	.521	.185	.034	.002
[π][p]	.011	.250	.512	.178	.037	.002
[π][p] - [π]	.001	.002	-.009	-.007	.003	.000

$_1[p]_{10}$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.044	.508	.395	.053	.000	.000
3	.007	.217	.570	.184	.022	.000
4	.000	.061	.500	.347	.092	.000
5	.000	.000	.454	.273	.182	.091
6	.000	.000	1.000	.000	.000	.000
[π]	.014	.248	.514	.188	.034	.002
[π][p]	.015	.249	.502	.182	.035	.003
[π][p] - [π]	.001	.001	-.012	-.006	.001	.001

TABLE VI.2.b January transition probability matrices for the maximum winds in the 10-15 km

layer over Cape Kennedy, Florida, 1956-1963. The speed class intervals are 0-20 mps, 21-40 mps,

..., 101-120 mps. The intervals are numbered 1 through 6. The prior (NOW) conditions are listed

vertically with the post (LATER) conditions listed horizontally. The matrices are identified as

$_\beta[p]_t$ where β is the month and t is the lag in time intervals of 12 hours, β = 1 (January);

t = 1, 2, ..., 10.

$$\left(_1[p]_1\right)^1$$

$$\left(_1[p]_1\right)^2$$

		LATER							LATER				
		1	2	3	4	5	6	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.000	.745	.246	.009	.000	.000	.000	.584	.374	.041	.001	.000
NOW	3	.000	.118	.764	.118	.000	.000	.000	.178	.651	.165	.006	.000
	4	.000	.000	.327	.622	.051	.000	.000	.039	.453	.444	.060	.005
	5	.000	.000	.000	.364	.545	.091	.000	.000	.210	.425	.316	.050
	6	.000	.000	1.000	.000	.000	.000	.000	.118	.764	.118	.000	.000

$(_1[p]_1)^3$

LATER

	NOW	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000
	2	.000	.479	.443	.075	.002	.000
	3	.000	.210	.595	.183	.012	.001
	4	.000	.082	.506	.351	.055	.005
	5	.000	.025	.349	.404	.194	.029
	6	.000	.178	.651	.165	.006	.000

$(_1[p]_1)^4$

LATER

1	2	3	4	5	6
.000	.000	.000	.000	.000	.000
.000	.409	.481	.104	.005	.000
.000	.226	.567	.190	.016	.001
.000	.121	.527	.299	.048	.005
.000	.060	.433	.363	.126	.018
.000	.210	.595	.183	.012	.001

$(_1[p]_1)^5$

LATER

	NOW	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000
	2	.000	.362	.503	.127	.008	.000
	3	.000	.235	.552	.193	.018	.001
	4	.000	.152	.535	.267	.041	.004
	5	.000	.096	.482	.323	.087	.011
	6	.000	.226	.567	.190	.016	.001

$(_1[p]_1)^6$

LATER

1	2	3	4	5	6
.000	.000	.000	.000	.000	.000
.000	.329	.515	.145	.011	.001
.000	.241	.544	.194	.020	.002
.000	.177	.538	.246	.036	.004
.000	.128	.509	.291	.064	.008
.000	.235	.552	.193	.018	.001

$$(_1[p]_1)^7$$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.306	.522	.158	.013	.001
3	.000	.243	.540	.194	.021	.002
4	.000	.195	.538	.231	.032	.003
5	.000	.156	.524	.265	.050	.006
6	.000	.241	.544	.194	.020	.002

$$(_1[p]_1)^8$$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.289	.527	.167	.015	.001
3	.000	.245	.538	.194	.021	.002
4	.000	.209	.538	.221	.029	.003
5	.000	.178	.531	.246	.041	.005
6	.000	.243	.540	.194	.021	.002

$$(_1[p]_1)^9$$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.278	.530	.174	.017	.001
3	.000	.246	.536	.194	.021	.002
4	.000	.219	.538	.213	.027	.003
5	.000	.195	.534	.232	.035	.004
6	.000	.245	.538	.194	.021	.002

$$(_1[p]_1)^{10}$$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.269	.531	.180	.018	.002
3	.000	.247	.536	.194	.022	.002
4	.000	.227	.537	.208	.026	.002
5	.000	.208	.536	.222	.031	.003
6	.000	.246	.536	.194	.021	.002

TABLE VI.2.c   January transition probability matrices for the maximum winds in the 10-15 km layer over Cape Kennedy, Florida, 1956-1963. The speed class intervals are 0-20 mps, 21-40 mps, ..., 101-120 mps. The intervals are numbered 1 through 6. The prior (NOW) conditions are listed vertically with the post (LATER) conditions listed horizontally. The matrices are identified as $_\beta[p]_t$ where β is the month and t is the lag in time intervals of 12 hours, $\beta = 1$ (January); t = 1, 2, ..., 10.

$$\left(_1[p]_2\right)^1$$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.666	.281	.053	.000	.000
3	.000	.158	.699	.132	.011	.000
4	.000	.020	.418	.491	.071	.000
5	.000	.000	.182	.545	.182	.091
6	.000	.000	.000	1.000	.000	.000

$$\left(_1[p]_2\right)^2$$

LATER

1	2	3	4	5	6
.000	.000	.000	.000	.000	.000
.000	.489	.406	.098	.007	.000
.000	.218	.590	.171	.019	.001
.000	.089	.516	.336	.052	.007
.000	.040	.388	.482	.074	.017
.000	.020	.418	.491	.071	.000

45

$$(_1[p]_2)^3$$

LATER

		1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000
	2	.000	.392	.463	.132	.013	.001
NOW	3	.000	.242	.549	.185	.022	.002
	4	.000	.148	.536	.273	.039	.005
	5	.000	.097	.497	.347	.052	.007
	6	.000	.089	.516	.336	.052	.006

$$(_1[p]_2)^4$$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	.337	.491	.154	.017	.001
3	.000	.252	.533	.190	.023	.002
4	.000	.188	.537	.239	.032	.004
5	.000	.150	.529	.276	.040	.005
6	.000	.148	.536	.273	.039	.005

$$(_1[p]_2)^5$$

LATER

		1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000
	2	.000	.305	.505	.169	.019	.002
NOW	3	.000	.256	.527	.192	.024	.002
	4	.000	.215	.534	.219	.029	.003
	5	.000	.189	.535	.240	.033	.004
	6	.000	.188	.537	.239	.032	.004

TABLE VI.2.d January difference matrices for the maximum winds in the 10-15 km layer over Cape

Kennedy, Florida, 1956-1963. The speed class intervals are 0-20 mps, 21-40 mps, ..., 101-120 mps.

The intervals are numbered 1 through 6. The prior (NOW) conditions are listed vertically with

the post (LATER) conditions listed horizontally. The matrices are identified as $\left(_\beta[p]_t\right)^n - _\beta[p]_t$

where β is the month, n is the power to which a matrix is raised, and t is the lag in time

intervals of 12 hours. $\beta = 1$ (January); $t = 1, 2, ..., 10$.

$$\left(_1[p]_1\right)^2 - _1[p]_2 \qquad\qquad \left(_1[p]_1\right)^3 - _1[p]_3$$

			LATER						LATER			
		1	2	3	4	5	6	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.000	-.082	.093	-.012	.001	.000	.000	-.126	.127	-.004	.002	.000
NOW	3	.000	.020	-.048	.033	-.005	.000	.000	.022	-.066	.043	.001	.001
	4	.000	.019	.035	-.047	-.011	.005	.000	.051	.047	-.066	-.027	-.005
	5	.000	.000	.028	-.120	.134	-.041	.000	.025	.167	-.323	.103	.029
	6	.000	.118	.764	-.882	.000	.000	.000	.178	.651	.165	-.994	.000

47

$$(_1[p]_1)^4 - {}_1[p]_4$$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	-.144	.113	.025	.005	.000
3	.000	.020	-.050	.028	.001	.001
4	.000	.070	.007	-.038	-.034	-.005
5	.000	.060	.342	-.455	.035	.018
6	.000	.210	.595	.183	-.988	.001

$$(_1[p]_1)^5 - {}_1[p]_5$$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	.000	-.199	.143	.048	.008	.000
3	-.004	.036	-.047	.017	-.004	.001
4	.000	.091	.004	-.070	-.020	-.006
5	.000	.096	.209	-.222	-.095	.011
6	.000	.226	.567	.190	-.984	.001

$$(_1[p]_1)^6 - {}_1[p]_6$$

LATER

NOW	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	-.009	-.179	.111	.066	.011	.001
3	-.004	.028	-.030	.003	.002	.002
4	.000	.095	.028	-.070	-.046	-.006
5	.000	.128	.145	-.163	-.118	.008
6	.000	.235	.552	.193	-.982	.001

$$(_1[p]_1)^7 - {}_1[p]_7$$

LATER

	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	-.026	-.177	.110	.079	.013	.001
3	.000	.022	-.022	.003	-.005	.002
4	.000	.093	.027	-.075	-.039	-.007
5	.000	.156	.160	-.189	-.132	.006
6	.000	.241	.544	.194	-.980	.002

$$\left({}_1[p]_1\right)^8 - {}_1[p]_8 \qquad\qquad \left({}_1[p]_1\right)^9 - {}_1[p]_9$$

		LATER							LATER			
	1	2	3	4	5	6	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	-.035	-.193	.097	.114	.015	.001	-.044	-.213	.126	.113	.017	.001
3 (NOW)	.000	.024	-.006	-.012	-.008	.002	.000	.029	-.030	.003	-.005	.002
4	.000	.117	.007	-.095	-.022	-.007	.000	.137	.007	-.103	-.034	-.007
5	.000	.087	.077	.064	-.232	.005	.000	.195	-.011	.141	-.329	.004
6	.000	.243	.540	.194	-.979	.002	.000	.245	.538	-.806	.021	.002

$$\left({}_1[p]_1\right)^{10} - {}_1[p]_{10}$$

		LATER				
	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000
2	-.044	-.239	.136	.127	.018	.002
3 (NOW)	-.007	.030	-.034	.010	.000	.002
4	.000	.166	.037	-.139	-.066	.002
5	.000	.208	.082	-.051	-.151	-.088
6	.000	.246	-.464	.194	.021	.002

TABLE VI.2.e   January difference matrices for the maximum winds in the 10-15 km layer over Cape Kennedy, Florida, 1956-1963. The speed class intervals are 0-20 mps, 21-40 mps, ..., 101-120 mps. The intervals are numbered 1 through 6. The prior (NOW) conditions are listed vertically with the post (LATER) conditions listed horizontally. The matrices are identified as $\left(_\beta[p]_t\right)^n - _\beta[p]_t$ where β is the month, n is the power to which a matrix is raised, and t is the lag in time intervals of 12 hours. $\beta = 1$ (January); $t = 1, 2, ..., 10$.

$$\left(_1[p]_2\right)^2 - {_1[p]_4} \qquad\qquad \left(_1[p]_2\right)^3 - {_1[p]_6}$$

		LATER						LATER				
		1	2	3	4	5	6	1	2	3	4	5	6
	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.000	-.064	.038	.019	.007	.000	-.009	-.116	.059	.052	.013	.000
NOW	3	.000	.012	-.027	.009	.004	.001	-.004	.029	-.025	-.006	.004	.001
	4	.000	.038	-.004	-.001	-.030	-.004	.000	.066	.026	-.043	-.043	-.005
	5	.000	.040	.297	-.336	-.017	.017	.000	.097	.133	-.107	-.130	.007
	6	.000	.020	.418	.491	-.929	.000	.000	.089	.516	.336	-.948	.006

six classes of equal intervals of 20 mps. A three dimensional, six-state, second order matrix can be represented by a two dimensional, 36-state, first order matrix where each of the 36 states is a combination of the original six states. Under this second order assumption the period-month wind data are processed by electronic computer to provide 36-state, lag $t = 1, 2, \ldots, 10$ matrices of prior against post observed conditions. The arrays contain frequencies of occurrence, relative frequencies, and empirical transition probabilities for each cell. The marginal or climatological vectors also are included.

For each matrix the following quantities are computed:

1. $[\pi] [p] - [\pi]$

2. H and R

3. $2n\bar{u}$ for independence between prior and post conditions

   (not corrected for zero frequencies).

The additional quantities $\left(_1[p]_1 \right)^t - \left(_1[p]_t \right)^1$ and $\left(_1[p]_2 \right)^t - \left(_1[p]_{2t} \right)^1$ are computed for the period-January matrices. All the calculations and matrices are held in file at ESSA-EDS-NWRC.

Manually tabulated $r = 2, 3, 4, 5$ dimensional, six-state, period-January matrices are tested for stationarity and specific order $m = 0, 1, 2, 3$. The test statistic used is the $2\hat{I}$ of V.C.2, where the p_a takes on the values appropriate for the order being tested. Normalized entropies also are computed. These manual calculating procedures are the same as those discussed previously in section VI.B.

VII. RESULTS

A. Stationarity

The metric chosen for comparison between the stationarity test $[\pi] [p] - [\pi] = p_{ij}$ and $[0] = q_{ij}$ is the largest state or norm

$$(_1[p]_2)^4 - {}_1[p]_8 \qquad\qquad (_1[p]_2)^5 - {}_1[p]_{10}$$

		LATER							LATER			
	1	2	3	4	5	6	1	2	3	4	5	6
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	-.035	-.145	.061	.101	.017	.001	-.044	-.203	.110	.116	.019	.002
NOW 3	.000	.031	-.011	-.016	-.006	.002	-.007	.039	-.043	.007	.002	.002
4	.000	.096	.006	-.077	-.019	-.006	.000	.154	.034	-.128	-.063	.003
5	.000	.059	.075	.094	-.233	.005	.000	.189	.081	-.033	-.149	-.087
6	.000	.148	.536	.273	-.961	.005	.000	.188	-.463	.239	.032	.004

difference $\max_{(ij)} |p_{ij} - q_{ij}|$. The arbitrarily selected criterion for rejecting the assumption of stationarity is a norm difference $\geq .030$. Of all the matrices tested the largest difference is .018, which is well below the criterion selected. It is assumed, therefore, that all the processes examined are stationary.

It is interesting to note, however, that the maximum norm difference occurs in October, while secondary maxima occur in the spring months of April and May. The operating processes in the summer months of July and August deviate the least from stationarity. These results reflect the abrupt change from summer to winter in October and the relatively gradual transition from winter to summer during the spring. The small norm differences in the summer are indicative of a constant weather regime.

B. Order

Baldwin's test statistic [5] for approximating the order of an operating system is calculated for each of the electronic computer tabulated matrices. The results of this test for first order matrices are illustrated by the graphs of the uncorrected $2n\bar{u}$ vs lag in Figures 7-9. The maximum orders m are determined from the point at which a curve intersects the horizontal line depicting the χ^2 rejection level. A summary of the approximate order for a given month and matrix class interval is shown in Table VII.1.

Clearly demonstrated from this table is the fact that as the class intervals become smaller, the persistence or history within the interval becomes less. This is not unexpected. Note that during the summer months the approximate order of Markovity for the 5 mps intervals is zero. This implies either that prediction within 5 mps
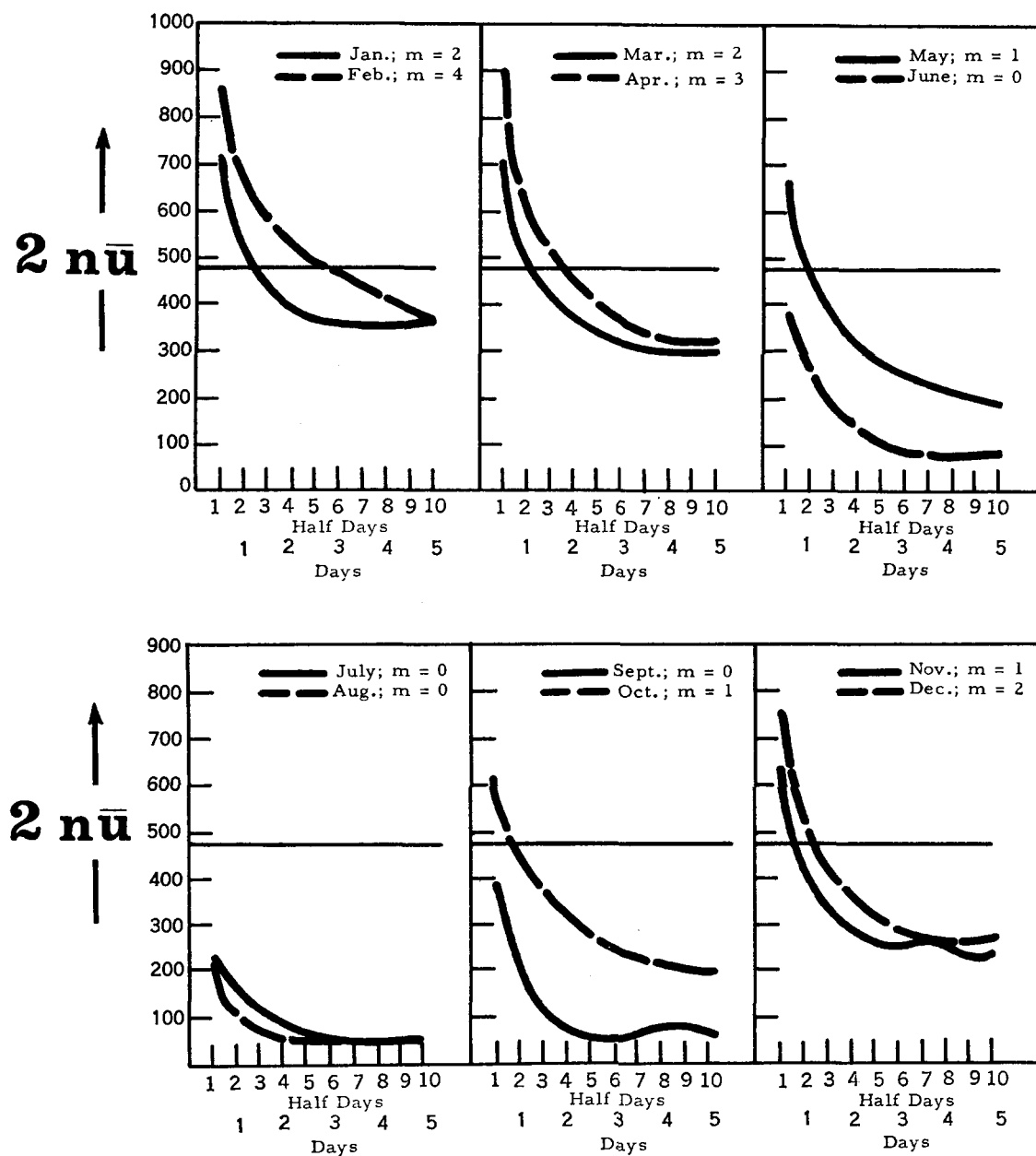
53
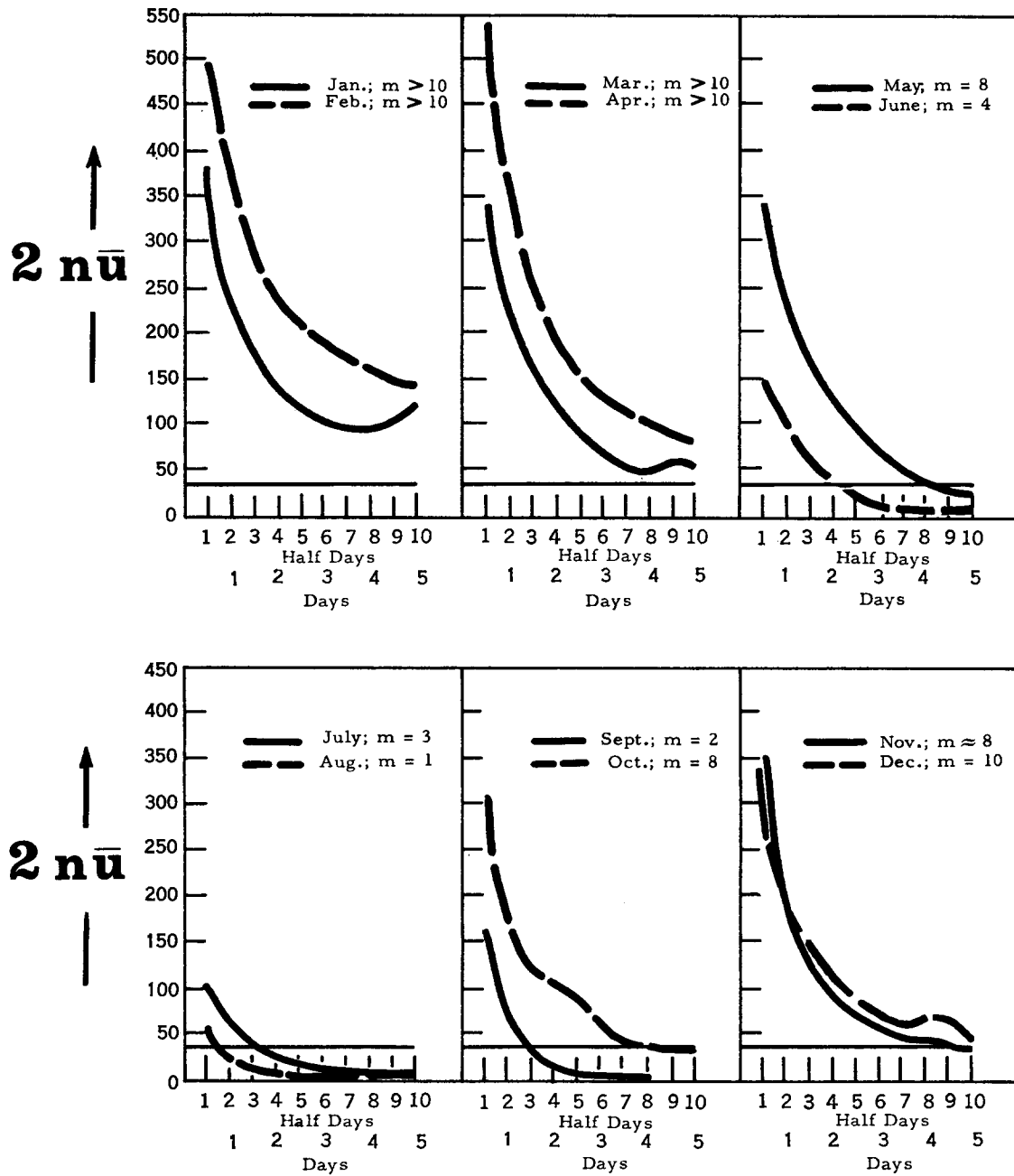
Fig. 7. Graphs of 2 n$\bar{u}$ vs. lag for Cape Kennedy, Florida, 10-15 km maximum wind matrices with class intervals of 0-20 mps, 21-40 mps, . . . , 101-120 mps for time intervals of 12 to 120 hours, 1/2 day to 5 days. Period of record 1956-1963. The uncorrected 2n$\bar{u}$ values are X^2 distributed with 25 degrees of freedom. The value of X^2 at the 0.95 rejection level, 34.4, is indicated by a horizontal line across each inset graph. " $\bar{u}$ " is the dependence capacitance.

54
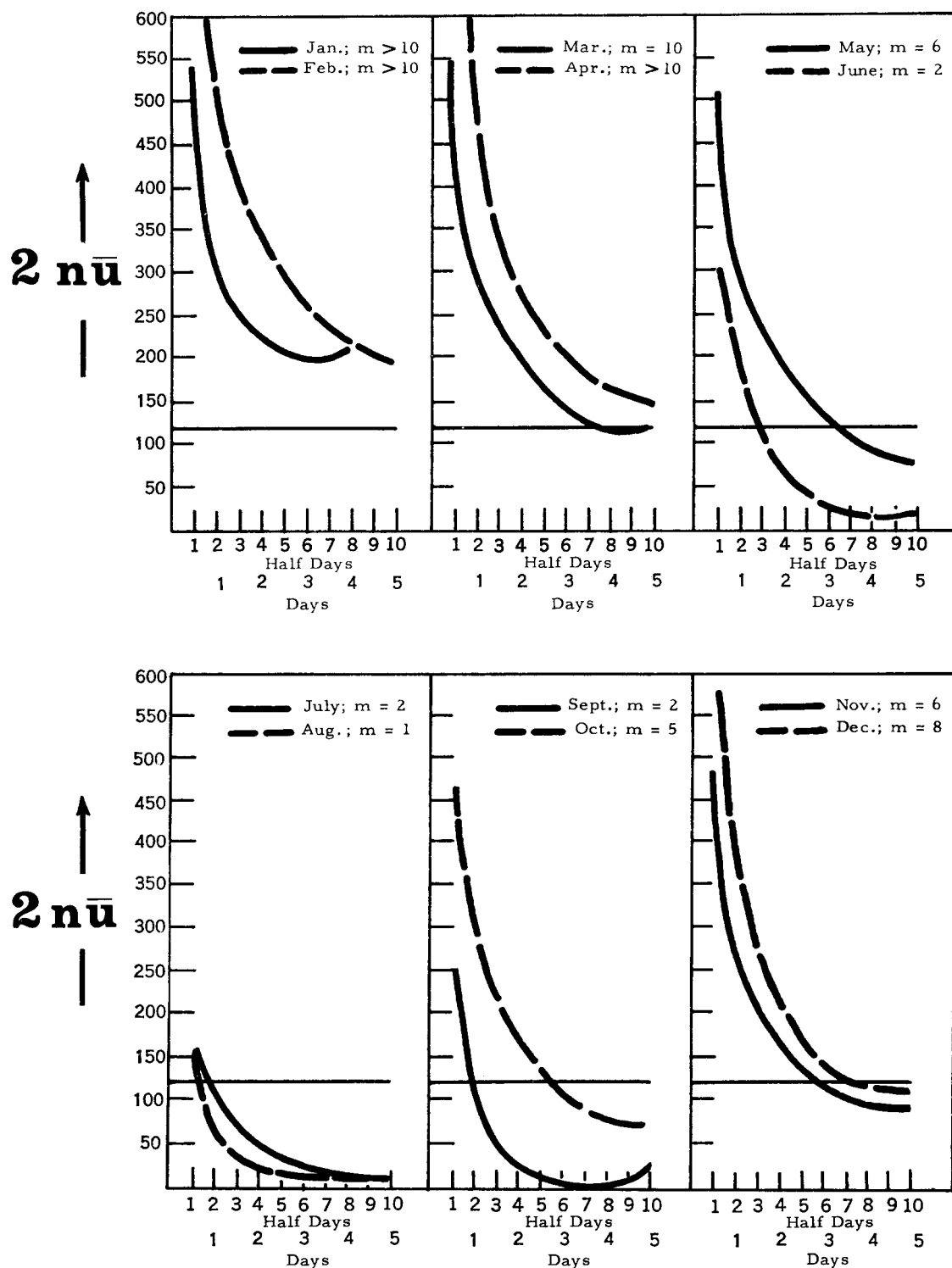
Fig. 8. Graphs of $2n\bar{u}$ vs. lag for Cape Kennedy, Florida, 10-15 km maximum wind matrices with class intervals of 0-10 mps, 11-20 mps, . . . , 101-110 mps for time intervals of 12 to 120 hours, 1/2 day to 5 days. Period of record 1956-1963. The uncorrected $2n\bar{u}$ values are χ^2 distributed with 100 degrees of freedom. The value of χ^2 at the 0.95 rejection level, 118.5, is indicated by a horizontal line across each inset graph. " $\bar{u}$ " is the dependence capacitance.

55

Fig. 9. Graphs of $2n\bar{u}$ vs. lag for Cape Kennedy, Florida, 10-15 km maximum wind matrices with class intervals of 0-5 mps, 6-10 mps,..., 106-110 mps for time intervals of 12 to 120 hours, 1/2 to 5 days. Period of record 1956-1963. The uncorrected $2n\bar{u}$ values are χ^2 distributed with 441 degrees of freedom. The value of χ^2 at the 0.95 rejection level, 476, is indicated by a horizontal line across each inset graph. " $\bar{u}$ " is the dependence capacitance.

TABLE VII.1 Approximate order of Markovity in the Cape Kennedy,
Florida, 10-15 km maximum wind matrices in terms of 1/2 day (12-
hour) periods by 20-, 10- and 5-mps class intervals. Period of
record 1956-1963. These orders have been determined from Figures 7-9.

Class Intervals
   (mps)

	J	F	M	A	M	J	J	A	S	O	N	D
20	10	10	10	10	8	4	3	1	2	8	8	10
10	10	10	10	10	6	2	2	1	2	5	6	8
5	2	4	2	3	1	0	0	0	0	1	1	2

is too rigid a requirement for the present system of observing and reporting, or that the transition processes have converged and that there is no useful information in going back more than one time period if one time period at all. In the latter case almost pure persistence could be used as a prediction. If any changes occur, they will be produced almost at random and a static prediction model would suffice. This does not deny, however, that there may be other avenues to explore. Other layers and other stations could be used in a spatial complex to provide more information.

The results of Baldwin's test [5] show that the operating processes are of minimum order in the summer and of maximum order in the winter. This means that there is more information to be gleaned from the history of the weather in the winter months than in the summer months. In other words the summer systems have nearly converged to climatology and the Markov models may not be effective in forecasting the weather during this time of year. On the other hand the Markov model prediction scheme may hold promise of some success during the winter months.

Monthly graphs of the uncorrected $2n\bar{u}$ vs lag for the second order six-state matrices are shown in Figure 10. The χ^2 rejection level based on $(36 - 1)^2$ degrees of freedom is not depicted because the number of degrees of freedom are invalid. The high frequencies of zero cells in these matrices make it possible to delete about 20 rows and columns from the winter arrays and about 30 rows and columns from the summer arrays. The resulting matrices would have the same computed $2n\bar{u}$ as the original ones, but the number of degrees of freedom would be greatly reduced.
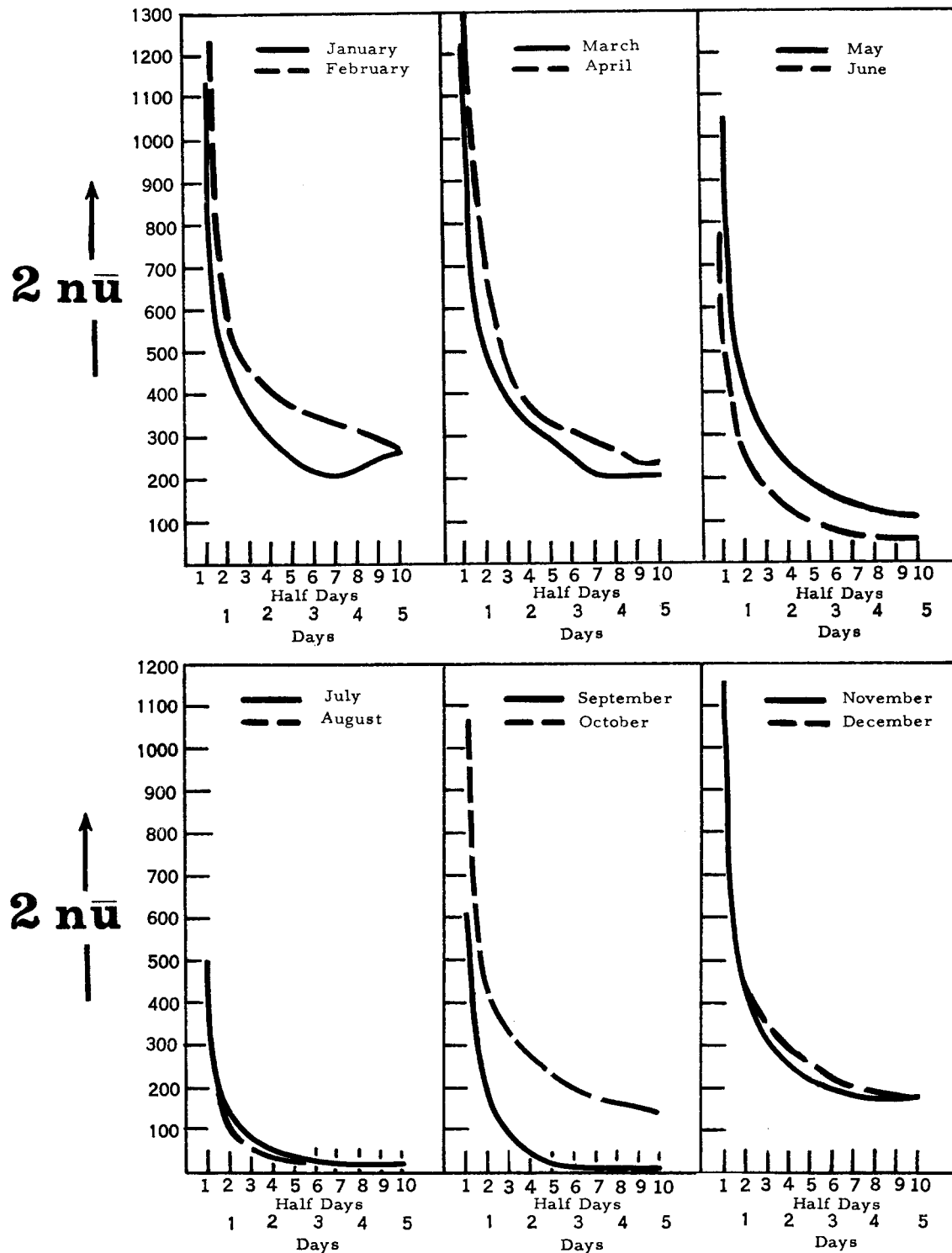
Fig. 10. Graphs of uncorrected $2n\bar{u}$ vs. lag for Cape Kennedy, Florida, 10-15 km maximum wind matrices with 36 classes $\{S_i \ S_j\}$ of interval S_i , S_j = 20 mps where i, j = 1, ..., 6. The time intervals are 12 to 120 hours, 1/2 day to 5 days. Period of Record 1956-1963. " $\bar{u}$ " is the dependence capacitance.

The tendency for some dependence capacitance curves to slope upwards towards five days implies that there may be some periodic function operating in the matrices and that the matrices are not actually stationary. However, this tendency is only slight and no further examination of the possible periodic feature is made here. Crutcher and Charles [13], though, previously noted four and five day periodicities in the wind at these levels in the Southeastern United States. Investigations of this feature will be made later.

A comparison between first and second order six-state matrices is made readily by examining normalized entropies. Figure 11 depicts the annual march of this quantity for the electronic computer tabulated matrices. The higher the normalized entropy, the more chaos or disorder that is inherent in the operating system, and the more information that may be extracted or utilized. A look at the two curves for six-state processes quickly reveals that in fall, winter and spring the second order matrices extract more information from the systems than do the first order matrices. In summer, however, the curves coincide and reach a minimum value. This means that most of the chaos is removed by the first order matrix, and that it will be very difficult to extract any more information from the operating systems.

The similar pattern of all four curves indicates that for the class intervals studied the winter months exhibit a high entropy and the summer months a low entropy. This implies that the history influence present in the winter may possibly be exploited to provide a workable Markov model, but that exploitation of the summer history influence will not appreciably improve a Markov model.
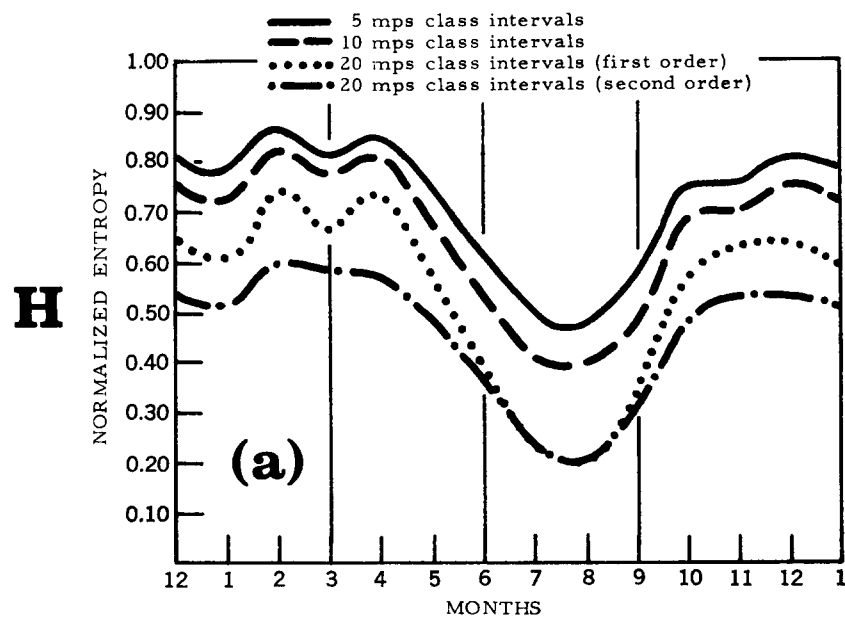
Fig. 11. Normalized entropy of Cape Kennedy, Florida,
10-15 km maximum wind transition matrices.
Period of record 1956-1963.

Because the above results indicate that a Markov scheme may
prove more fruitful for the winter months than for the summer months,
January data is studied in more detail than the data from other months.
The corrected $2\hat{I}$ of Kullback et al [40] for testing order is computed
for 2-, 3- and 6-state transition matrices. The null hypotheses tested
are that the matrices are of order $m = 0, 1, 2, 3$. The 0.95 con-
fidence level is chosen as the basis for non-rejection of the hypothe-
ses. The degrees of freedom with which to enter the χ^2 table are
adjusted to compensate for estimates of zero of theoretical proba-
bilities. The results of these tests are presented in Table VII.2.
It is seen that although the test of Baldwin revealed possible rela-
tively high orders, the data can be modeled by low order Markov chains.

In order to check these results the normalized entropies R of
each s-state, m-th order process are tabulated. An arbitrarily
selected change of $\geq$ 10 percent between R_m and R_{m+1} is considered
to be significant. Under this criterion it is found that all four
January systems are second order. Use of higher order models does
not produce a significant gain of information.

Under the assumption of stationarity the 6 x 6, 11 x 11 and
22 x 22 matrices are tested for first order Markovity by comparing
$\left(_\beta[p]_1\right)^t$ and $\left(_\beta[p]_2\right)^t$ with $\left(_\beta[p]_t\right)^1$ and $\left(_\beta[p]_{2t}\right)^1$, respectively.
The metric selected for comparison is the norm distance. Interpre-
tation of this quantity, however, is difficult. With only one
observation in a cell, obviously a difference of transition proba-
bilities could be 1.00. Since the norm value is dependent upon the
number of observations used to determine an empirical cell transition
probability, the problem arises as to what constitutes the minimum

TABLE VII.2 Order of Markovity as determined by the test of Kullback et al [40] in the Cape Kennedy, Florida, 10-15 km maximum January wind matrices in terms of 1/2 day (12-hour) periods. Period of record: 1956-1963.

Number of States s	Class Intervals	Order
2	$0 \leq s_1 < 43$ mps $\leq s_2$	2
2	$0 \leq s_1 < 70$ mps $\leq s_2$	3
3	$0 \leq s_1 < 43$ mps $\leq s_2 < 70$ mps $\leq s_3$	1
6	$0 \leq s_1 < 20$ mps, ..., 100 mps $\leq s_6 < 120$ mps	2

number of cell observations that would make the norm value meaningful. Further research towards the application of binomial or multinomial probabilities to this problem is suggested.

A subjective evaluation of the norm differences is possible. For the three s-state systems the highest values occur in the winter and the lowest values in the summer. This infers that a first order model would fit the summer data better than the winter data.

C. Verification of the Prediction Schemes

The electronic computer tabulated transitional and conditional probability matrices are used as prediction schemes to make forecasts out to three days or six time intervals during the winter season of December, January and February and during the summer season of June, July and August. Thirty days during each season are randomly selected from the three-year period 1964-1966. The 10-15 km maximum wind observations on these days determine the initial conditions from which predictions are made. In some instances less than 180 forecasts are made for a season. This occurs either because the conditions during the test period exceed the limits of the predictive scheme or because the predictive scheme will allow equiprobable outcome events.

For each of the s-state systems, forecasts are made using first order conditional probability matrices $\left(_\beta [p]_t\right)^1$ and Markov transitional probability matrices $\left(_\beta [p]_1\right)^t$. Two predictions of persistence and one of climatology also are made. The first persistence forecast verifies if the later observation falls in the same class as the initial observation, where the class interval is determined by the s-states. The second persistence forecast verifies if the later

64

observation falls in the same class interval as the initial observation, where the class interval is such that the initial observation is in the middle of the interval. The state with the highest frequency of observations from 1956-1963 serves as the climatological forecast. In addition to the above schemes second order conditional probability matrices are used to forecast six-state conditions.

The results of these forecasts are depicted in Table VII.3. It is readily apparent that the prediction accuracy decreases as the class interval is made smaller. This is not unexpected since an increase in the number of classes is concomitant with a departure from the perfect but meaningless forecast implied by a one-state system. It is noted, though, that the relative frequencies are very low in the 5 mps class interval systems. The inference is that the requirement for predicting in a high-precision, 22-state system is too rigid for the present practices of observing upper level winds.

The accuracy of all the forecast schemes is greater in the summer than in the winter because the warm weather winds are much less variable than the winter winds. This low variability also explains the high verification scores of persistence. It will be difficult for any prediction scheme to be better than persistence during the summer season.

A Markov first order prediction scheme works best in the 11-state system. It provides the greatest accuracy of any winter forecast and also does well in the summer for a process with 10 mps class intervals. In a system with more states the precision requirements are too stringent to obtain good forecasts, and in a system with fewer states

TABLE VII.3 Relative frequencies of successful forecasts of the maximum wind in the 10-15 km

   layer over Cape Kennedy, Florida. The prediction schemes are based on 1956-1963

   data. Six forecasts of 12, ..., 72 hours are made for each of 30 randomly selected

   days in the winter and summer seasons during the test period 1964-1966.

	WINTER (December, January, February)			SUMMER (June, July, August)		
	6-state 20 mps	11-state 10 mps	22-state 5 mps	6-state 20 mps	11-state 10 mps	22-state 5 mps
Persistence 1*	.550	.294	.144	.774	.483	.278
Persistence 2**	.600	.294	.156	.800	.544	.283
Climatology	.439	.389	.100	.572	.439	.267
1st Order Conditional Probability Matrices	.617	$.303^1$	$.157^2$	.656	$.441^4$	$.277^5$
1st Order Markov Probability Matrices	.589	.417	.206	.617	$.485^6$	$.283^7$
2nd Order Conditional Probability Matrices	$.619^3$	--	--	$.689^8$	--	--

*   Class interval determined by the s-states.
** Class interval determined such that the initial observation is in the middle of the interval.
1   Based on 178 forecasts.
2   Based on 172 forecasts.
3   Based on 160 forecasts.
4   Based on 170 forecasts.
5   Based on 166 forecasts.
6   Based on 171 forecasts.

7   Based on 173 forecasts.
8   Based on 177 forecasts.

most of the variability of the observations is within a class and cannot be detected by the Markov scheme.

Manually tabulated first, second, and third order one-step transition matrices for January are used to make 12-hour forecasts on 40 randomly selected days from the four Januaries in 1964-1967. The forecasts are compared with those of climatology and persistence. The four systems shown in Table VII.2 are evaluated.

The results of the 12-hour forecasts are given in Table VII.4. In all four systems the climatological forecast is the worst. The Markov scheme predictions coincide with those of persistence in all but the 6-state system. Although forecasts of transitions other than persistence are allowed in the 2- and 3-state systems, instances where they can be made are limited to the lowest frequency classes. In the 6-state system the third order matrix provides the best forecasts of the three Markov schemes.

Markov models will produce results equal to or better than persistence. In the one case where persistence shows a higher score the comparable Markov model is not shown. It would be necessary to construct a new model based on the same interval used for each persistence prediction. This unequal comparison should also be kept in mind while assessing the results previously shown in Table VII.3.

VIII. CONCLUSIONS

The results of the previous section can be construed to indicate that a forecast of the 10-15 km maximum winds over Cape Kennedy, Florida, based on persistence is not significantly different from a forecast based on a more sophisticated Markov scheme. This inference, however, is limited in light of several considerations of the study.

TABLE VII.4 Relative frequencies of successful 12-hour forecasts of the January maximum wind in the 10-15 km layer over Cape Kennedy, Florida. The prediction schemes are based on 1956-1963 data. One forecast is made for each of 40 randomly selected days from the four Januaries in 1964-1967.

	2-state[a]	2-state[b]	3-state[c]	6-state[d]
Persistence 1*	.850	.975	.825	.725
Persistence 2**	--	--	--	.850
Climatology	.450	.925	.375	.350
1st Order Markov Probability Matrices	.850	.975	.825	.718[1]
2nd Order Markov Probability Matrices	.850	.975	.825	.711[2]
3rd Order Markov Probability Matrices	.850	.975	.825	.737[2]

*   Class interval determined by the s-states

** Class interval determined such that the initial observation is in the middle of the interval

a $0 \leq s_1 < 43$ mps $\leq s_2$

b $0 \leq s_1 < 70$ mps $\leq s_2$

c $0 \leq s_1 < 43$ mps $\leq s_2 < 70$ mps $\leq s_3$

d $0 \leq s_1 < 20$ mps, ..., 100 mps $\leq s_6 < 120$ mps

1   Based on 39 forecasts.

2   Based on 38 forecasts.

Some considerations of the limitations of the data are in order.
The forecast schemes are based on data from an eight-year period, or
essentially from a relatively small sample of eight. Since the wind
distribution is continuous, it is expected that over a long period
of time all cells of a contingency table should be filled, even though
the matrix represents a discrete distribution. The discreteness is
only artificially induced by arbitrarily setting class intervals into
which the observations are forced to fall. Because of the small
sample of eight that is used in this study, most of the matrices con-
tain cells of zero frequencies.

The problems encountered as a result of the zero cell frequencies
are numerous. The empirical relative frequency matrices are used
as estimates of the true theoretical transition probability matrices.
Most of the matrix testing procedures, however, are valid only for
theoretical probabilities greater than zero. The correction for zero
frequencies in an observed distribution that is being tested for fit
by a theoretical distribution is empirical and therefore subject to
error. The evaluation of the degrees of freedom of a system becomes
extremely tedious when there are zero cells in a matrix because each
component of the system must be treated separately. Finally, forecast
schemes that allow a zero probability of occurrence for some classes
are not at all satisfying.

One obvious method of alleviating some of the above problems
is to use a larger sample. Twelve years of serially complete wind
data will soon be available at ESSA-EDS-NWRC. It is questionable,
though, whether the 50 percent increase in data over the present study
will be enough to stabilize the systems and to eliminate the cell

frequencies of zero. Another method is to embed the observed frequency distributions in matrices of ones before determining relative frequencies. Although this technique is less satisfying theoretically than intuitively, the embedding process approximates a continuous distribution and the problem of zero cell frequencies would be eliminated. This feature will be studied.

Periodicities within the data may be hindering the effectiveness of the Markov prediction schemes. It is suggested that spectra or harmonic analyses be made on the data to determine if significant cyclical influences are present. If so, these cycles should be removed from the data series, and the residual series should be studied. The forecast scheme then will consist of a component from the significant cycles and a component from the residual series. This feature will be studied.

Another problem concerns the reliability of the empirical transition probability matrices. Confidence bounds need to be established on all of the probabilities. Perhaps binomial or multinomial probability theory can be applied further in the solution of this problem. Future investigation also should be made to determine the critical levels for accepting or not accepting the hypotheses tested by the metric tests.

The capacity of the present and planned electronic computers at ESSA-EDS-NWRC severely restricts the study of processes of order greater than one or two for more than a few states. The size of matrices and the number of combinations of possible events increase very rapidly as the number of states and/or orders increases. This expansion quickly causes the limits of the computer to be exceeded and makes the task of manually processing the data monumental. It is recommended,

however, that higher order processes should be studied for at least the 6- and 11-state systems.

In view of the aforementioned problems this study can be considered as a first step towards the prediction by the use of Markov techniques of the maximum winds in the 10-15 km layer over Cape Kennedy, Florida. The results herein are in part encouraging and in part discouraging. At least the technique does as well as persistence and holds promise for better prediction. Only after much more investigation will the merits of the Markov technique as applied to the maximum winds be able to be assessed fully.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

1. Allen, R. G., C. A. Champion, F. E. Courtney and M. J. Andre, "Objective Empirical Approaches to Extended Forecasting and Short Term Weather Cycles," ER-5718 (1962), Lockheed Nuclear Products, Lockheed-Georgia Co., Marietta, Georgia, pp. 1-145.

2. Anderson, A. C., "A Study of the Accuracy of Winds Derived from Transosonde Trajectories," NRL Memorandum Rpt. No. 498 (1955), Naval Research Laboratory, Aerology Branch, Atmosphere and Astrophysics Division, Washington, D. C., pp. 1-10.

3. Andre, Milo J., Use of Climatic Transition Probabilities in Problems of Applied Climatology and Extended Forecasting, presented at the 192nd National Meeting of the American Meteorological Society, held jointly with the Section of Meteorology, American Geophysical Union, Washington, D. C., 1961.

4. Baldwin, J. G., "On the Application of the Theory of Non-stationary Markov Chains to Short Term Prediction of Climatological Variables," Tech. Rpt. SU-192 (1964), Research Triangle Institute, Durham, North Carolina, pp. 1-16.

5. Baldwin, J. G., "The Application of M-th Order Markov Chains to Short Term Prediction of Climatological Variables," Final Rpt. SU-226 (1965), Research Triangle Institute, Durham, North Carolina, pp. 1-14.

6. Bartlett, M. S., An Introduction to Stochastic Processes with Special Reference to Methods and Applications, Cambridge Univ. Press, Cambridge, 1966.

7. Billingsley, Patrick, "Statistical Methods in Markov Chains," Ann. Math. Statist., Vol. 32 (1961), pp. 12-40.

8. Billingsley, Patrick, Statistical Inference for Markov Processes, Institute of Mathematical Statistics, Univ. of Chicago Statistical Research Monographs, Univ. of Chicago Press, Chicago, 1961.

9. Caskey, James E., Jr., "A Markov Chain Model for the Probability of Precipitation Occurrence in Intervals of Various Length," Mon. Wea. Rev., Vol. 91 (1963), pp. 298-301.

10. Chung, K. L., Markov Chains with Stationary Transition Probabilities, Springer-Verlag, Berlin, 1960.

11. Crutcher, Harold L., "Statistical Prediction of Upper Air Winds for the Purpose of Synthesizing Missing Observations," unpublished manuscript (1957), National Weather Records Center, Asheville, North Carolina.

12. Crutcher, Harold L., personal statement (1963).

13. Crutcher, H. L. and B. N. Charles, "Spectrum Analysis of Upper Winds," unpublished manuscript (1960), National Weather Records Center, Asheville, North Carolina.

14. Crutcher, H. L. and R. Durham, "Prediction of Wind Profiles at Cape Kennedy, Florida with Emphasis on Multiple Regression Techniques," unpublished manuscript (1965), National Weather Records Center, Asheville, North Carolina.

15. Crutcher, H. L. and F. Orovitz, "Prediction of Wind Profiles at Cape Kennedy, Florida with Markov (Transition probability) Matrix Techniques," unpublished manuscript (1965), National Weather Records Center, Asheville, North Carolina.

16. Crutcher, H. L. and F. T. Quinlan, "Prediction (Estimation) of
    Cape Kennedy, Florida Wind Speed Profile Maxima," unpublished
    manuscript (1964), National Weather Records Center, Asheville,
    North Carolina.

17. Dynkin, E. B., Die Grundlagen der Theorie der Markoffschen Prozesse,
    Springer-Verlag, Berlin, 1961.

18. Feinstein, A., Foundations of Information Theory, McGraw-Hill, New
    York, 1958.

19. Feller, William, An Introduction to Probability Theory and Its
    Applications, 2nd ed., Vol. 1, John Wiley and Sons, New York, 1957.

20. Feyerherm, A. M. and L. D. Bark, "Statistical Methods for Persistent
    Precipitation Patterns," J. Appl. Meteor., Vol. 4 (1965), pp. 320-
    328.

21. Fisher, R. A. "On the Mathematical Foundations of Theoretical
    Statistics," Phil. Trans. A, Vol. 222 (1922), pp. 309-368.

22. Fisher, R. A., "On a Distribution Yielding the Error Functions of
    Several Well Known Statistics," Proc. Int. Math. Congress, Toronto
    (1924), pp. 805-813.

23. Gabriel, J. E. and R. Bellucci, "Time Variations of Winds Aloft,"
    J. Meteor., Vol. 8 (1951), pp. 422-423.

24. Gabriel, K. R. and J. Neumann, "A Markov Chain Model for Daily
    Rainfall Occurrence at Tel Aviv," Quart. J. Roy. Meteor. Soc.,
    Vol. 88 (1962), pp. 90-95.

25. Getman, F. H. and F. Daniels, Outlines of Theoretical Chemistry,
    John Wiley and Sons, New York, 1931.

26. Godske, C. L., "Contributions to Statistical Meteorology, I,"
    Goephysica Norvegica, Vol. XXIV (1962), pp. 161-210.

27. Godske, C. L. "Statistics of Meteorological Variables," Final
    Rpt. AF61(052)-416 (1965), Univ. of Bergen, Bergen, Norway,
    pp. 1-115.

28. Godske, C. L., "Some Studies in Statistical Meteorology,"
    Scientific Rpt. No. 1, AF61(052)-760 (1966), Univ. of Bergen,
    Bergen, Norway, pp. 1-28.

29. Gringorten, Irving I., "A Stochastic Model of the Frequency and
    Duration of Weather Events," J. Appl. Meteor., Vol. 5 (1966),
    pp. 606-624.

30. Herdan, G., Language as Choice and Chance, P. Noordhoff N. V.,
    Groningen, 1956.

31. Keeping, E. S., Introduction to Statistical Inference, D. Van
    Nostrand Co., Princeton, 1962.

32. Kemeny, J. G. and J. L. Snell, Finite Markov Chains, D. Van
    Nostrand Co., Princeton, 1960.

33. Kolmogorov, A., "Anfangsgrunde der Markoffschen Ketten mit Unendlich
    Vielen Moglichen Zustanden," Rec. Math., Moscou (Mat. Sbornik),
    N. S., Vol. 1 (1936), pp. 607-610.

34. Koopman, B. O., "A Generalization of Poisson's Distribution for
    Markoff Chains," Proc. Nat. Acad. Sci. U. S. A., Vol. 36 (1950),
    pp. 202-207.

35. Koopman, B. O., "A Law of Small Numbers in Markoff Chains," Trans.
    Amer. Math. Soc., Vol. 70(1951), pp. 277-290.

36. Ku, H. H., "A Note on Contingency Tables Involving Zero Frequencies and the 2I Test," *Technometrics*, Vol. 5 (1963), pp. 398-400.

37. Ku, H. H., personal statement (1968).

38. Kullback, S., *Information Theory and Statistics*, John Wiley and Sons, New York, 1959.

39. Kullback, S., M. Kupperman and H. H. Ku, "An Application of Information Theory to the Analysis of Contingency Tables, with a Table of $2n \ln n$, $n = 1(1)10,000$," *J. Res., Nat. Bur. Stand., B*, Vol. 66B (1962), pp. 217-243.

40. Kullback, S., M. Kupperman and H. H. Ku, "Tests for Contingency Tables and Markov Chains," *Technometrics*, Vol. 4 (1962), pp. 573-608.

41. Kullback, S. and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 79-86.

42. Kupperman, Morton, "On Comparing Two Observed Frequency Counts," *Applied Statistics*, Vol. 9 (1960), pp. 37-42.

43. Masuyama, Motosaburo, "Tables of n, $\log_e n$, $n \log_e n$ and $n(\log_e n)^2$ for $n = 1$ Through 500 with Applications," *Rep. Statist. Appl. Res.-Un. Jap. Sci. Engrs.*, Vol. 7 (1960), pp. 56-64.

44. Neyman, J. and E. S. Pearson, "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika*, Vol. 20A (1928), pp. 175-240, 263-294.

45. Parzen, Emanuel, *Stochastic Processes*, Holden-Day, San Francisco, 1962.